VOL. XXVI MAY, 1935 NO. 5

# The Journal of Educational Psychology

*Devoted Primarily to the Scientific Study of Problems of Learning and Teaching*

## CONTENTS

*Empirical Evaluation of Three Personality Adjustment Inventories* 321
CARNEY LANDIS, JOSEPH ZUBIN AND SIEGFRIED E. KATZ

*Predictive Value of the Vocational Interest Test* . . . . . 331
EDWARD K. STRONG, JR.

*Types of Assumptions in Educational Research*. . . . . 350
DOUGLAS E. SCATES

*Comparative Pupil Gains under Individual Conference and Classroom Instruction*. . . . . 367
GEORGE W. HARTMANN

*A Formula for Finding the Average Correlation of Any One Variable with the (n − 1) Other Variables without Solving Any of the Individual Correlations* . . . . . 373
HAROLD A. EDGERTON

*The Influence of College Training upon Success after College as Measured by Judges' Estimates*. . . . . 377
W. R. VAN VOORHIS AND A. C. MILLER

*The Measurable Effects of Zoology on Psychology Grades*. . . . . 384
A. R. GILLILAND

*A New Method of Increasing the Reliability of the True-false Test* 388
C. IVER HOVLAND AND J. C. EBERHART

*Note on Scoring Tests of Multiple Weighted Items*. . . . . 395
VIRGINIA I. ZERILLI

*Book Reviews* . . . . . 398

**$6.00 per Year · Published Monthly September to May**

**WARWICK & YORK, INC.**

**BALTIMORE, MD.**

**Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.**

# THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

*Established* 1910

EDITORS

JACK W. DUNLAP
Fordham University

HAROLD E. JONES
University of California

PERCIVAL M. SYMON
Teachers College,
Columbia University

H. E. BUCHHOLZ, *Managing Editor*

---

THE price of the Journal is $6.00 a year in the United States; $6.40 foreign countries. Part-year subscriptions are 90 cents for each number ordered. Back volumes are $7.00 each; back issues are $1.10 each except when more than five years old, and then $1.20 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Unsolicited manuscripts should be accompanied with return postage. Manuscripts, books and other materials for review, and correspondence regarding editorial and business matters should be addressed to the Journal.

WARWICK AND YORK · *Publishers* · BALTIMORE, MD.

An Experimental Study of Certain Factors Affecting Transfer of Training in Arithmetic

BY JAMES ROBERT OVERMAN

An investigation of the effect of instruction on three types of examples in two-place addition upon the pupils' ability to handle closely related types in addition and subtraction, and to determine whether the amount of transfer is a function of the method of teaching. The problem was to ascertain if the amount of such transfer can be increased by helping the pupils consciously to generalize the process and to formulate, from the type taught, a general method of procedure applicable to related types.

$3.00 *plus* 12¢ *postage*

WARWICK AND YORK, *Publishers*, BALTIMORE, MD.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXVI | May, 1935 | Number 5

# EMPIRICAL EVALUATION OF THREE PERSONALITY ADJUSTMENT INVENTORIES

CARNEY LANDIS, JOSEPH ZUBIN,[1] AND SIEGFRIED E. KATZ

New York State Psychiatric Institute and Hospital

Since Woodworth's[2] questionnaire was devised and published, many revisions have appeared adding improvements to his original inventory but usually retaining its general form and some of its items.

In the attempt to validate these tests, research workers have made use of the entire gamut of social deviants, from the problem child through the delinquent and criminal to psychoneurotic and psychotic individuals. In the present study a psychopathological group and a non-psychopathological group were used as criterion groups in the validation of the questionnaire items. Three recent personality tests of the Woodworth Psychoneurotic Inventory pattern were used.

## PURPOSE

The purpose of this study was to determine whether the three tests purporting to measure personality adjustment differentiate normally adjusted individuals from those who are so maladjusted that they require institutional care. Mental Hospital patients were given the Bernreuter[3] Personality Inventory, the Page[4] Questionnaire of Schizophrenic Traits and the Maller[5] Character Sketches.

[1] Read in part before the meeting of the New York Branch of the American Psychological Association, New York University, 1934.

[2] Published in Hollingworth, H. L.: *Psychology of Functional Neuroses.* New York, Appleton, 1920, pp. 120–126.

[3] Bernreuter, R. G.: *The Personality Inventory.* Stanford Univer. Press, Stanford Univer., Calif., 1931.

[4] Page, J., Landis, C., and Katz, S. E.: "Schizophrenic Traits in the Functional Psychoses and in Normal Individuals." *Amer. Jour. Psychiat.*, Vol. XIII, 1934, pp. 1213–1225.

[5] Maller, J. B.: *Character Sketches.* Bureau of Publications, Teachers College, Columbia University, New York, 1932.

For each patient a normal individual of the same chronological age, intelligence, academic status, economic background (occupation), racial extraction, religion and nativity was selected.[1]

## RESULTS

### *Subjects*

For the group of individuals used in the study of the Maller Character Sketches the median chronological age was 26.0 years; the median highest school grade attained was 10.7 (second year of high school); the median average age-grade of leaving school was 16.4 years. The inter-quartile range for these three figures was 7.9, 7.9, and 1.7 respectively. The populations used in the Bernreuter and Page studies did not differ significantly from this one in any of these respects. With respect to the age-grade status our population samples, both normal and abnormal, seem to be about that of the population in general.

Of the entire population 40.2 per cent were males and 59.8 per cent females. The distribution by racial origin was as follows: 27.7 per cent were Jews, 7.1 per cent Italians, 7.1 per cent German, 5.4 per cent Irish, about 30 per cent were of mixed racial origin and the remainder came from various European countries. Sixteen per cent were professional people themselves or came from families of professional status, 35.7 per cent were retail merchandisers, 27 per cent were skilled laborers and the remainder were scattered among the various other occupational categories. *Each patient was matched with a normal individual who was similar to the patient in all the above characteristics.* The patients comprised four clinical categories. Table I shows the distribution of the subjects by categories. (It should be noted that on the Page Questionnaire, the matching was not done individually but for the group as a whole.)

The normal individuals were obtained from College classes, settlement houses, employment bureaus, etc. The tests were administered individually or in small groups. The abnormal individuals were obtained at the New York State Psychiatric Institute and the Utica

[1] We wish to thank Dr. M. R. Trabue and the staff of the Adjustment Service Bureau for their kindness in permitting the use of their collected data; also Dr. J. B. Maller and Miss Margaret Quayle for the use of data collected by them; also Mrs. M. B. Creelman, Miss J. Henning, Mr. J. Page and Mr. L. DeAngelis, for part of the data collected from patients.

TABLE I.—DISTRIBUTION OF SUBJECTS BY DIAGNOSIS AND TESTS

| Category | Bernreuter | Character sketches | Page questionnaire |
|---|---|---|---|
| Normal | 125 | 107 | 240 |
| Schizophrenic | 44 | 37 | 40 |
| Manic depressive | 34 | 30 | 30 |
| Psychoneurotic | 28 | 25 | 18 |
| Organic (paretic) | 19 | 15 | 11 |

State Hospital. They were first-admission patients and represent a fair sampling of the more cooperative of the mental hospital patients.

### BERNREUTER PERSONALITY INVENTORY

In the study of the critical ratios of the differences between the entire normal and entire abnormal groups by sex, we found that for the Bernreuter B1-N Scale a critical ratio of 0.04 existed between the males, 0.76 between the females and a critical ratio for the total group of 0.57. With respect to the B2-S Scale the critical ratio for males was 1.78, for females 2.65 and for the total group 3.14. For Scale B4-D the critical ratio for males was 0.33, for females 1.80, and for the total group 1.49. (A critical ratio must be over 3.0 to be significant.) On the basis of these ratios the B1-N Scale which purports to measure neurotic tendency, may be said to show no significant differences between the normal and abnormal groups.

The comparisons thus far have dealt with the entire normal and abnormal groups. A comparison of the critical ratios of the differences between the individuals belonging to the various categories of psychiatric diagnosis and their normal controls is shown in Table II.

TABLE II.—BERNREUTER PERSONALITY INVENTORY
Critical Ratios for Normal-abnormal Comparisons by Categories

| Group | Scales | | | |
|---|---|---|---|---|
| | N | B1-N | B2-S | B4-D |
| Schizophrenic | 44 | 0.16 | 1.79 | 1.30 |
| Manic depressive | 34 | 1.59 | 2.15 | 1.15 |
| Psychoneurotic | 28 | −2.05 | 2.13 | 1.43 |
| Organic | 19 | 1.79 | −1.15 | −0.45 |

When each category is compared with its own control group in Table II, such sex differences as were mentioned above disappear. On the basis of this evidence it can be stated that the Bernreuter Personality Inventory does not differentiate between the abnormal individuals included in this study and the normal individuals with whom they were matched. This finding is in essential accord with that reported by Landis and Katz,[1] who found that of two hundred twenty psychopathic patients, the percentage falling in the highest and lowest quartiles of the range of scores for normals was as follows: B1-N, upper quartile, nineteen, lower thirty-seven; B2-S, upper, thirty-seven and lower, twenty-one; B4-D, upper thirty-nine and lower, eighteen. If the test were to meet the specification stated by its author, all or nearly all members of a neurotic population should fall into the upper (neurotic) quartile. Apparently Scale B1-N does not differentiate between normals and abnormals.

### THE CHARACTER SKETCHES

The "Character Sketches" have two distinguishing features. First, each item is presented twice, once as a positive item and again as a negative item. Thus, "Is very popular at a party" is counterbalanced by "Is not popular at a party." This permits a check on the consistency of the response. The second feature is that the items are arranged in some logical order and are grouped under the following categories: "Habit Pattern," "Self Control," "Social Adjustment," "Personal Adjustment," "Mental Health," and "Readiness to Confide." Each one of the subgroups of items yields a score for the category it purports to measure.

Table III gives the means and standard deviations of the normal and abnormal groups on the Character Sketches. The difference is sixteen points (critical ratio, 3.70), which is statistically significant. In the case of this test too, the differences become less pronounced when each category is compared with its own control group.

When the comparison was made by psychiatric diagnostic categories, the group of thirty-seven schizophrenic patients gave a critical ratio with their normal control of 0.87; thirty manic depressive patients a critical ratio of 0.92; twenty-five psychoneurotic patients a critical

---

[1] Landis, C. and Katz, S. E.: "The validity of certain questions which purport to measure neurotic tendencies." *Jour. Appl. Psychol.*, Vol. XVIII, 1934, pp. 343–356.

TABLE III.—CHARACTER SKETCHES
Total Scores of the Psychopathic and Normal Groups

| | | Total score | |
|---|---|---|---|
| | N | Ave. | *O* |
| Normal | 107 | 142.57 | 26.79 |
| Abnormal | 107 | 126.58 | 35.77 |
| Difference | ... | 15.99 | 4.32 |
| Critical ratio | ... | ..... | 3.70 |

ratio of 3.86; and fifteen organic (paretic) patients a critical ratio of −1.00. Only one of these ratios, namely, the psychoneurotic, may be said to be significant.

### PAGE QUESTIONNAIRE OF SCHIZOPHRENIC TRAITS

The Page Questionnaire of Schizophrenic Traits consisted of statements concerning the possession of traits that are generally accepted by psychiatrists as symptomatic of schizophrenia.

Table IV gives the means and standard deviations on the Page Questionnaire of Schizophrenic Traits. The difference between the normal and abnormal groups is not significant, as was pointed out previously by Page, Landis and Katz.

TABLE IV.—PAGE QUESTIONNAIRE
Total Score of the Psychopathic and Normal Groups

| | N | Ave. | *O* |
|---|---|---|---|
| Normal | 240 | 22.4 | 5.89 |
| Abnormal | 99 | 22.8 | 7.53 |
| Difference | ... | 0.4 | 0.84 |
| Critical ratio | ... | .... | 0.48 |

### ITEM ANALYSIS

Reviewing the findings which we have reported above with respect to these three tests, it is apparent that even on two such extreme groups as psychopathic patients and normals there is considerable overlap in total score on the tests. There is one obvious reason for this lack of differentiation, namely, the method of validation of the tests, which was primarily logical rather than empirical.

The construction of the key for scoring the answers to the items on these questionnaires has been based on the consensus of opinion of trained psychologists and psychiatrists. This crystallized opinion served as the basis for judging whether a given response to an item should be considered neurotic or not. Thus because most of the judges agree that excessive "day dreaming" should be regarded as an indication of neurotic tendency, an answer of "yes" to the question: "Do you day-dream a lot?" is marked as neurotic. Similarly an answer of "yes" to—"Do ideas often run through your head so that you cannot sleep?" is marked as neurotic. On the basis of a key prepared in this manner, the total score for each individual in all the items is obtained, and the criterion of internal consistency[1] is applied to determine whether a given item is "good" or not. If the item differentiates between the high scorers—those individuals who obtained a high total score on all the items combined—and the low scorers, the item is considered good and retained. If the item fails to differentiate between the high scorers and low scorers, it is eliminated.

Without in any way questioning the validity of the observation that "day-dreaming" and "sleeplessness" are neurotic symptoms, an attempt was made to find out whether normal people actually differ from abnormal people in response to such items.

The responses to each item on the three questionnaires were tabulated and a card prepared giving the frequency of each response by categories. Table V shows such a card for item twenty-one of the Bernreuter: "Do ideas often run through your head so that you cannot sleep?"

The independence value (value expected by chance) of the frequency for each cell was computed and the $\chi^2$ value of the discrepancy between the independence value and the obtained value was determined. Since the table is $3 \times 6$, the number of degrees of freedom is $2 \times 5 = 10$ and the value of $\chi^2$ that may arise by chance only once in one hundred times for this type of table is 23.209. Hence, whenever $\chi^2$ summed for the entire table is 23.09 or more,[2] the table cannot be explained by the hypothesis that the answers are random, but only by the hypothesis that certain categories are associated with certain responses.

---

[1] For a discussion of the methods of internal consistency see: Zubin, J.: "The method of internal consistency for selecting test items." *Jour. Educat. Psychol.*, Vol. XXV, 1934, pp. 345–356.

[2] Fisher, R. A.: *Statistical Methods for Research Workers.* Oliver and Boyd, London, 1930. Sec. 22, p. 91—Partition of $x^2$ into its components.

TABLE V.—DISTRIBUTION OF RESPONSES TO BERNREUTER ITEM #21 FOR NORMAL AND PSYCHIATRIC GROUPS

| Group | Yes | | | No | | | ? | | | $x^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | *O* | *E* | $x^2$ | *O* | *E* | $x^2$ | *O* | *E* | $x^2$ | |
| Normal..... | 84 | 70.7 | 2.5 | 32 | 44.3 | 3.4 | 9 | 10.0 | .09 | 6.3 |
| Schizophrenic[1]..... | 16 | 23.8 | 2.5 | 20 | 14.9 | 1.8 | 6 | 3.3 | 2.1 | 6.4 |
| Manic depressive.. | 13 | 19.2 | 2.0 | 19 | 12.1 | 4.0 | 2 | 2.7 | 0.2 | 6.2 |
| Psychoneurotic.... | 20 | 15.8 | 1.1 | 6 | 9.9 | 1.6 | 2 | 2.2 | 0.0 | 2.6 |
| Organic..... | 7 | 10.7 | 1.3 | 11 | 6.7 | 2.7 | 1 | 1.5 | 0.2 | 4.2 |
| Total..... | 140 | .... | ... | 88 | .... | ... | 20 | .... | .... | 25.7 |

*O* = Obtained; *E* = Expected.

[1] Two cases failed to respond to this item.

In order to determine in which category the association occurs, the value of $\chi^2$ for each row must be investigated. Since each row consists of three cells, there are two degrees of freedom and $\chi^2$ must equal at least 9.2 before the presence of any association may be inferred.

No single category in Table VIII yields a $\chi^2$ which is significant, but the $\chi^2$ for normal—abnormal differentiation is quite significant. It can be shown readily that the value of $\chi^2$ for each category when pitted against all the remaining categories is equal to the value of $\chi^2$ obtained for that category in this table multiplied by $\frac{N}{N-f}$ where $f$ is the frequency of the category. For the normal group, $\frac{N}{N-f} = 2$. Hence, the obtained value of $\chi^2$ of 6.3 must be doubled and becomes equal to 12.6, which value could occur by chance less than one time in one hundred.

To confirm this, an additional analysis was made of each diagnostic category in which only those individuals that belonged to the given category were compared with their matched controls. Each category gave rise to a $3 \times 2$ table and the results generally were in agreement with the findings of the first analysis. Wherever there was any disagreement, the results of the second analysis were given preference.

A comparison of the logically expected abnormal response with the empirically determined abnormal response revealed some unexpected reversals. Contrary to logical expectation, normals say significantly more often than abnormals that they

1. Day-dream frequently.

2. Cross the street to avoid meeting some person.
3. Have ideas run through their heads so that they cannot sleep.
4. Think they could become so absorbed in creative work that they would not notice a lack of intimate friends.
5. Usually enjoy spending an evening alone.
6. Find it difficult to get rid of a salesman.
7. Prefer to be alone at times of emotional stress.
8. Usually work better when praised.
9. Do not give money to beggars.

The Bernreuter Scale credits all the above items as neurotic symptoms. The above is not to be taken as proof that the symptoms involved are not truly neurotic symptoms. It may, however, be concluded that on such self-rating scales as these questionnaires present, the maladjusted individuals tend to answer contrary to logical expectations. Whether this is due to lack of insight or to desire to dissimulate is difficult to say. The fact remains that normal individuals do give the responses listed above.

On the basis of this analysis the differentiation power of the items in each of the tests was determined.

TABLE VI.—NUMBER OF DIAGNOSTIC ITEMS BY CATEGORIES

| Test | Total abnormal | Categories | | | |
|---|---|---|---|---|---|
| | | Schizophrenic | Manic depressive | Psychoneurotic | Paretic |
| Bernreuter inventory..... | 17 | 5 | 2 | 2 | 4 |
| Maller character sketches | | | | | |
| Part I..... | 18 | 7 | 1 | 15 | 2 |
| Part II..... | 21 | 9 | 1 | 21 | 7 |
| Page questionnaire..... | 15 | 8 | 6 | 7 | 0 |

If the items that were found to be diagnostic (Table VI) were scored in accordance with a key based on this empirical analysis, the Bernreuter test would differentiate between normal and abnormal individuals, the Character Sketches would differentiate between the total normal and abnormal groups and especially between psychoneurotic individuals and normals, and the Page Questionnaire, like the Bernreuter, would differentiate between normals and abnormals.

All the items that proved to be diagnostic were put together into a new theoretical scale. Three such theoretical scales were drawn up—

Abnormal, Psychoneurotic and Schizophrenic. The Abnormal Scale consisted of forty-four items that differentiated the total normal from the total abnormal group. Theoretical scores on this Abnormal Scale were obtained for each individual by giving a credit of one unit for each abnormal response to the thirty-two items. Similar scores were obtained for the Schizophrenic and Psychoneurotic groups.

The average score of each group on each of these scales was determined and the percentage that this score was of the total maximum score was computed. Table VII gives the results.

TABLE VII.—PERCENTAGE OF MAXIMUM OBTAINED ON EACH SCALE BY CATEGORIES

| Group | Scales | | |
|---|---|---|---|
| | Schizophrenic | Psychoneurotic | Abnormal |
| Normal | 11.2 | 16.0 | 22.9 |
| Schizophrenic | 31.7 | 38.5 | 40.6 |
| Psychoneurotic | 21.7 | 61.5 | 42.7 |
| Abnormal | 30.9 | 38.4 | 42.0 |
| Number of item | 20 | 17 | 32 |

On the schizophrenic scale the normals obtained on the average 11.2 per cent of the total number of schizophrenic traits. The schizophrenic group itself obtained 31.7 per cent of the maximum number of schizophrenic items. The remainder of the table is to be read similarly. These scales deal only with the responses positively diagnostic of each of the categories, that is, with those responses whose frequency in the normal group was greater than that expected by chance (and consequently, whose frequency in the abnormal group was less than that expected by chance, or vice versa). A response whose frequency in the normal group was significantly less than expected (and whose frequency in the abnormal group was consequently greater than expected) is called a negatively diagnostic item. An item which did not differentiate between the normal and abnormal was called a neutral item. For one scale—the abnormal scale, the average number of diagnostically positive, negative and neutral items was determined for both categories. Table VIII gives these averages.

The normal group possesses 8.56 abnormal traits but 17.82 or twice as many normal traits, which gives the normal group a net positive score of 9.26. The abnormal group possesses some normal traits, but an equal number of abnormal traits, giving it a net score of

TABLE VIII.—AVERAGE NUMBER OF DIAGNOSTICALLY POSITIVE, NEGATIVE AND NEUTRAL RESPONSES FOR THE NORMAL AND ABNORMAL GROUPS

| | Positive | Negative | Neutral | Tentative score |
|---|---|---|---|---|
| Normal..... | 17.82 | 8.56 | 5.62 | 9.26 |
| Abnormal..... | 12.89 | 12.73 | 6.38 | 0.16 |

about zero. There is no great difference in the number of neutral traits for the two groups.

Apparently, the three tests failed to differentiate between normals and abnormals because they were not validated on the basis of their differentiating power between normals and abnormals but on another basis, namely, the criterion of internal consistency applied to logically selected items. Whether or not a scale built on such a criterion is useful in the measurement of adjustment is debatable. It has been shown in this article that such tests do not measure the type of maladjustment which characterizes mental patients when those patients are compared with matched normals.

## SUMMARY

To summarize the results, two of the tests do not differentiate between normal and abnormal individuals. On the third test, the Character Sketches, although a statistical difference is obtained when the *psychoneurotic* group is compared with its control group, overlap between the groups is still present. If the tests actually measure adjustment, two extreme groups, such as the abnormal and normal groups under comparison, should give very little or no overlap.

An analysis of the items showed that the reason for the lack of differentiation can be attributed to the *logical* rather than empirical evaluation of the items. An empirical evaluation of these items yielded a tentative scale which seems to differentiate normals from abnormals, as well as the various psychiatric categories.

This analysis, together with the previous studies of Page, Landis and Katz, and Landis and Katz, indicates clearly the unsatisfactory nature of the personality inventories at present available. Until we have personality inventories which measure or at least indicate those factors in the life of the individual which the inventory purports to measure, psychologists can do no other than regard these questionnaires as instruments of research not yet ready for general application.

# PREDICTIVE VALUE OF THE VOCATIONAL INTEREST TEST

EDWARD K. STRONG, JR.

Stanford University

The ten-year follow-up of fourteen-year-old pupils by Professor Thorndike and his colleagues indicates that tests of intellect, clerical ability, and mechanical adroitness at age fourteen, and measures of school adjustment up to that age, are predictive of later educational progress, but scarcely at all predictive of success in clerical occupations or mechanical pursuits.[1] This conclusion means that vocational guidance in elementary and secondary schools is probably fraught with greater difficulties than anyone has realized. Lorge reports that "the predictions are better for age twenty to twenty-two than for age eighteen to twenty" and expresses the hope "that the prediction for later years may be better still." It is certainly desirable that a further follow-up may be made upon this group, for the unstable nature of occupational choice in youth suggests that the determinations made at age fourteen may prove to be more useful after twenty years than has been found to be the case after only ten years.

The investigation reported here differs from the above in two main respects: First, the measurements pertain to interests and not to intelligence, scholarship, and specific abilities; second, the subjects are college seniors, not grammar school pupils. These subjects would all measure quite high on any intelligence test or scholarship basis, since they were college seniors when first encountered.

In March 1927 all the senior men at Stanford University were requested to fill out the Vocational Interest Test and a questionnaire regarding their occupational plans. Slightly over two-thirds complied, a total of two hundred eighty-seven men. In March 1932 these men were asked to fill out a second interest blank, two hundred twenty-three responding. The data which follow are based on these two hundred twenty-three cases. In January 1928, January 1929, and March 1932, the group reported their occupational progress, giving thereby a remarkably full record of the jobs held and the occupational plans in mind on four different dates. In many cases very enlightening comments were included.

[1] Lorge, I: "The Prediction of Vocational Success." *The Personnel Journal*, December, 1933, pp. 189–196.

To what extent do college graduates maintain the occupational plans they had as seniors in college? The answer to this question is considered first as it has some bearing upon the principal question to be answered in this article, which concerns the relationship between occupational interest ratings in 1927 and occupational choice in 1932.

### CONSTANCY OF OCCUPATIONAL CHOICE OF SENIORS

Forty-eight per cent of the group have not changed their occupational choice, nine per cent have made what appears to be a rather slight change, twenty-two per cent have made a real change, and twenty-one per cent have changed from "don't know" to a specific choice (Table I). The writer has spent considerable time studying all the data regarding each man before making the above classification. In view of the great complexity of the data it is improbable that any two persons would exactly agree upon the proper grouping, or even that the writer could exactly duplicate the present grouping on another occasion. What has been done is sufficiently accurate, however, so that one may conclude that one half of these men have continued their occupational program for five years, that one fourth have changed their plans quite radically, and one fourth have selected an occupation whereas five years ago they reported they did not know what they were going to do. Not much greater precision in this regard can be obtained until an adequate classification of occupations is developed. As it is, we do not know whether shifting from psychiatrist to psychologist is a real change or not. On the other hand, all would agree that shifting from law to farming or from law to art represents very distinct changes in occupational choice.

The one hundred six seniors who have not changed their occupation are reported in Table II. Only those who have selected an occupation for which the vocational interest test can be scored are reported in the body of the table, the remainder being given in a footnote. The seventy seniors who have changed their occupations are reported in Table III and those who have changed from "don't know" to a definite occupation in Table IV.

Is there any evidence that men entering the professions are more likely to continue therein after graduation than men entering business? If we assume all the changes included in Tables III and IV are significant, then it appears that ninety-two per cent of those selecting chemistry in 1927 have adhered to that program, eighty-five per cent of those selecting medicine have continued therein, and similarly

seventy-five per cent of lawyers, sixty-one per cent of engineers, forty per cent of school men, and only twenty-nine per cent of business men. Evidently those planning to enter a profession are more likely to continue therein than those selecting business. This fact would be more strikingly shown if business were subdivided into three or more subdivisions, for then it would appear that relatively few have continued in the same sub-division of business. Two reasons may be advanced for this situation. First, undergraduates at Stanford University obtain more intensive professional and semi-professional courses in engineering, law, medicine, chemistry, and the like than they do in business. Because of this condition those who are uninterested or unfit are weeded out and the remainder are more likely to continue after graduation in one of the professions than in business. Second, it appears to the writer that men entering the professions do so by actual choice whereas to a considerable degree men entering business do so because they are not interested in any particular profession.

### ADEQUACY OF VOCATIONAL INTEREST TEST

The requirements for an adequate vocational guidance test are threefold: Reliability, permanence, and validity. A person's score should agree very closely with the score he would obtain if he took the test a second time, both within a short interval of time (reliability) and after a considerable period of time (permanence). His score should furthermore be indicative of the occupation he should enter (validity). The average reliability of twenty-one scales of the Strong Vocational Interest Blank is .89. Permanence of interests for two hundred twenty-three college men over a five-year period is .75 (or .84 when corrected for attenuation).[1] The measurement of permanence in terms of a correlation coefficient indicates the extent to which the individuals maintain the same rank order on both occasions. But such a measurement does not reveal whether the scores change, becoming higher or lower for the entire group, during the designated interval of time. Actually, occupational interest scores average one-fourth of a quartile higher in 1932 than in 1927. One might suppose that if a man's score should rise in engineering it would fall in an occupation which correlates negatively with engineering, as for example life insurance selling (correlation of $-.72$). This does happen to some

[1] Strong, E. K., Jr.: "Permanence of Vocational Interests." *Jour. Educ. Psychol.*, May, 1934, p. 339.

extent. But another factor primarily affects the situation, namely age. The occupational scales are based on the interests of adult men, known to be successfully employed in the respective occupations. Younger men are penalized on all the scales because their interests differ somewhat from older men.[1] The differences between scores in 1927 and 1932 are accordingly very largely a measure of the changes in interest between twenty-two and twenty-seven years of age.

Determination of the validity of a vocational test is fraught with many difficulties. What should be the criterion? At first thought "final vocational choice" appears to be the only ultimate criterion in guidance. But after further consideration it is apparent that one can not assume that every man eventually enters the occupation for which he is best fitted. If this were true there would be no great need for vocational tests, although even then the use of tests might save some men several years of fumbling around. Because final occupational choice can not be accepted as a perfect criterion, it necessarily follows that a vocational test which correlates perfectly with final occupational choice is as faulty as the present system of finding one's livelihood. Whether the two should correlate .30, .50, or .70 is beyond our powers today to determine. It is, however, quite likely that this correlation should be much higher in the case of college graduates than for the great majority, since the former have had a superior opportunity to choose their careers.

Apparently the only feasible way to validate a guidance test is to prove first that those who are satisfactorily adjusted in an occupation may be differentiated by the test from those who are not satisfactorily adjusted, including also those not engaged in the occupation; and second to show that young people who obtain scores comparable to the satisfactorily adjusted group enter that occupation in reasonable degree and those who do not obtain such scores enter other occupations. We cannot expect to obtain correlations of 1.00 in this respect because economic conditions do not permit complete freedom of choice. We know that the vocational interest test differentiates between men in different occupations. We are here concerned with the second phase of the matter, that is, the extent to which college graduates enter the occupation in which they score highest on the interest test.

---

[1] Strong, E. K., Jr.: *Change of Interests with Age.* 1931.

———: "Interest Maturity." *Personnel Journal,* Vol. XII, No. 2, August, 1933.

As evidence of validity, the vocational interest scores of two hundred twenty-three college seniors will be considered in the light of the following four propositions which should be true if the interest test is significant in this connection:

1. Men continuing in an occupation obtain a higher interest score in it than in any other occupation;
2. Men continuing in an occupation obtain a higher interest score in it than men entering some other occupation;
3. Men continuing in an occupation obtain higher scores in it than men who change from that occupation to some other;
4. Men changing from other occupations to occupation *X* score higher in *X* prior to the change than they did in other occupations.

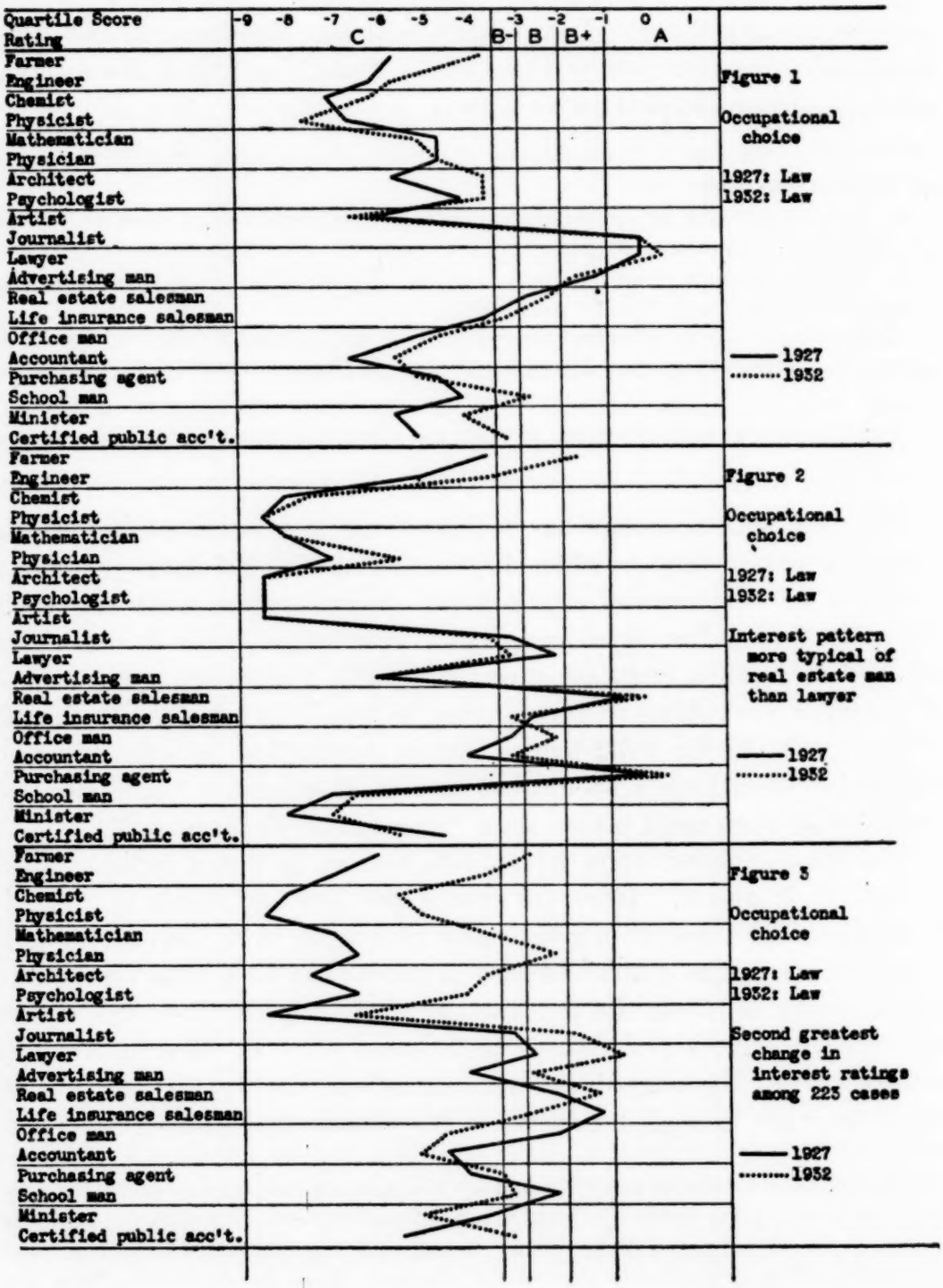
## RELATION OF VOCATIONAL INTEREST RATINGS TO OCCUPATIONAL PROGRAM

Occupational program has reference to the occupational *choice* expressed by our subjects when they were seniors (1927 data) but it has reference to "occupation engaged in" in connection with the 1932 follow-up. In a few cases exceptions were made when it was clear that the occupation engaged in was definitely temporary in nature, as for example work done to earn money to *return* to school. It should be noted that in the 1932 figures higher correlation between interest ratings and vocational choice could have been secured if "occupational choice" had been used rather than "occupation engaged in."

The kind of data upon which this section is based is illustrated in Figs. 1, 2, and 3. Here are shown the interest scores of twenty occupations for three young men now practising law; the scores have been converted into quartile scores in order that all will be on a comparable basis as far as the five ratings of *A*, *B*+, *B*, *B*−, and *C* are concerned.[1] The three figures give some idea of how individual records vary. Upon graduation from college the young man whose record appears in Fig. 1 entered a law school and since graduation

[1] The converted scores are obtained by subtracting the median from the raw score and dividing the result by the quartile value of the criterion group. Quartile scores above −1 are rated *A*; between −1 and −2, they are rated *B*+, etc. On a few scales, due to markedly skewed distribution, there are deviations from the above procedure (see the published scales).

A rating of "*A*" means the person has the interests of the occupation, a rating of "*C*" means he does not have such interests. Ratings of *B*+, *B*, and *B*− are intermediate, meaning he probably has those interests but one can not be so sure of it as in the case of the *A* rating.


Quartile Score
Rating
-9 -8 -7 -6 -5 -4 -3 -2 -1 0 1
C
B-
B
B+
A
Farmer
Engineer
Chemist
Physicist
Mathematician
Physician
Architect
Psychologist
Artist
Journalist
Lawyer
Advertising man
Real estate salesman
Life insurance salesman
Office man
Accountant
Purchasing agent
School man
Minister
Certified public acc't.
Figure 1
Occupational choice
1927: Law
1932: Law
1927
1932
Figure 2
Occupational choice
1927: Law
1932: Law
Interest pattern more typical of real estate man than lawyer
1927
1932
Figure 3
Occupational choice
1927: Law
1932: Law
Second greatest change in interest ratings among 223 cases
1927
1932

from it has practised law for two and a half years. His interest ratings agree with this program, since he rates higher in law than in any other occupation. As to the *A* rating in journalism, it may be pointed out that the correlation between the law and journalism scales is .826, which becomes .90 when corrected for attenuation. (See also Fig. 5 for additional evidence that this interest in journalism is typically associated with interest in law.) It is therefore appropriate that a young lawyer should also have a high rating in journalism. Figure 2 gives a different constellation of interests, in which purchasing and real estate rate distinctly higher than law in 1927. In 1932 these two rate slightly higher still and law lower. This is seemingly the picture of a real estate man rather than a lawyer. Figure 3 shows the interests of a life insurance man in 1927, with lesser interests in real estate, office work, and teaching. The interests of this young man changed quite noticeably in five years, for in 1932 law stands first, with real estate and journalism second and third, and life insurance a poor sixth.

It should be noted here that these examples are not used as proof of the validity of the scale, but merely as illustrations of the sort of data being considered. The last two cases challenge our interest; they indicate that some time must elapse before final conclusions are drawn concerning the diagnostic value of the test. If the final outcome indicates that the original prediction of the test was correct, it will constitute very important evidence of the great value of measured interests in a vocational guidance program.

## CONSIDERATION OF THE FOUR PROPOSITIONS

All of our data sustain our first proposition, that men continuing in an occupation obtain higher interest scores in that occupation than in any other. Table V gives such data for five occupational groups, all the groups that we have containing ten or more cases.[1] In each group the highest score is obtained in the occupation in which the group is engaged. This fact can be seen at a glance by reference to Figs. 5 and 6 which give the data for twenty-seven engineers and fifteen lawyers for twenty occupations and not merely for nine occupations as in the table. According to these scores, the average engineer obtained in 1927 a $B+$ rating in engineering and chemistry, a $B$ rating in farming and purchasing, and a $B-$ rating in office work.

[1] The table contains all available data from the two hundred twenty-three seniors and a few other comparable cases from our files.

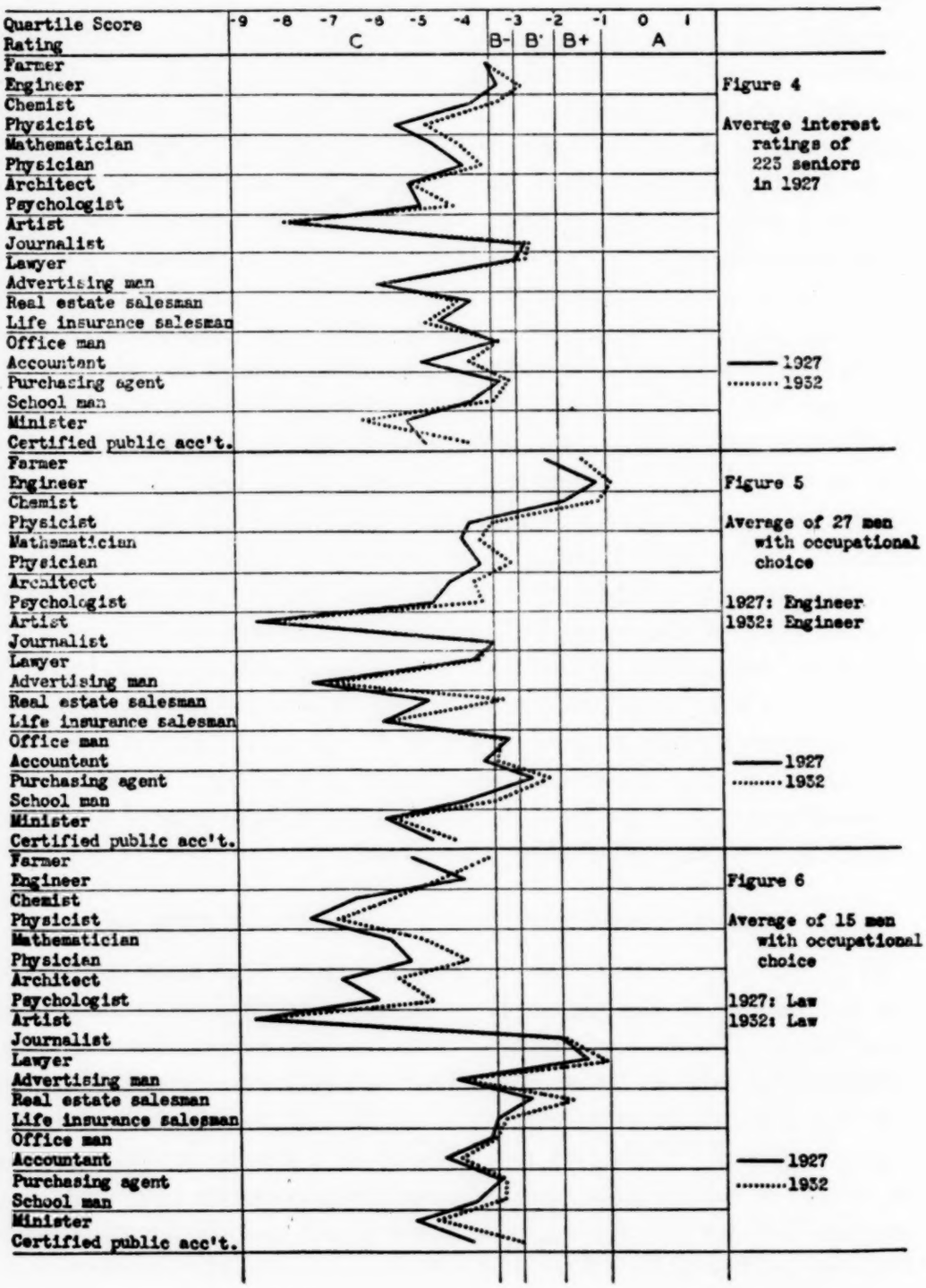

Quartile Score
Rating
-9 -8 -7 -6 -5 -4
C
-3 -2 -1 0 1
B- B' B+ A
Farmer
Engineer
Chemist
Physicist
Mathematician
Physician
Architect
Psychologist
Artist
Journalist
Lawyer
Advertising man
Real estate salesman
Life insurance salesman
Office man
Accountant
Purchasing agent
School man
Minister
Certified public acc't.
Figure 4
Average interest ratings of 225 seniors in 1927
——1927
.....1932
Figure 5
Average of 27 men with occupational choice
1927: Engineer
1932: Engineer
Figure 6
Average of 15 men with occupational choice
1927: Law
1932: Law

TABLE I.—PERMANENCE OF OCCUPATIONAL CHOICE OF COLLEGE SENIORS FOR FIVE YEARS

| | Number | Per cent |
|---|---|---|
| No change in occupational choice | 106 | 47.6 |
| Slight change in occupational choice | 21 | 9.4 |
| Distinct change in occupational choice | 49 | 22.0 |
| Change from "don't know" to definite choice | 47 | 21.0 |
| Total | 223 | 100.0 |

TABLE II.—ONE HUNDRED SIX SENIORS WHOSE OCCUPATIONAL CHOICE WAS THE SAME IN 1927 AND 1932; ALSO MEDIAN RANKING OF INTEREST SCORES FOR OCCUPATION GIVEN, WHEREVER AVAILABLE

| No. | Occupational choice[1] | Median ranking of interest test in | | |
|---|---|---|---|---|
| | | 1927 | 1932 | Difference |
| 27 | Engineer | 1.96 | 2.03 | − .07 |
| 12 | Chemist | 1.05 | 1.18 | − .13 |
| 17 | Physician | 3.00 | 1.69 | 1.31 |
| 1 | Psychologist | 2.50 | 1.00 | 1.50 |
| 1 | Architect | 1.00 | 1.00 | 0 |
| 15 | Lawyer | 1.50 | 1.21 | .29 |
| 2 | Writer | 1.50 | 1.25 | .25 |
| 1 | Life insurance salesman | 2.50 | 2.50 | 0 |
| 1 | Minister[2] | 11.50 | 12.50 | −1.00 |
| 4 | School man | 2.00 | 2.75 | − .75 |
| 1 | Certified public accountant | 2.00 | 1.50 | .50 |
| 24 | Miscellaneous[3] | | | |
| | Median of eighty-two cases | 1.78 | 1.65 | .13 |
| | Quartile of distribution | 1.00 | 1.14 | |

[1] As the two hundred twenty-three men had chosen only twenty of the occupations for which the interest test may be scored, only those twenty interest scales have been considered in this study. They are: Engineer, chemist, physicist, mathematician, physician, architect, psychologist, farmer, artist, journalist (writer), lawyer, advertising man, salesman of life insurance, salesman of real estate, minister, school man, purchasing agent, office worker, accountant, and certified public accountant.

[2] In explanation of the low rankings of this minister is this comment by him: "When I say my occupation was 'more or less thrust upon' me, I mean that when I was ready for college I had the choice of no further education or studying for the ministry, which called for a college and seminary education. Desiring further education, I chose the latter."

[3] The twenty-four miscellaneous cases include: four in finance, three geologists, three zoologists, three entomologists, two in business in general, two in marketing, two "don't know" and one each in diplomatic service, sales engineering, professor of history, physical education teacher and technician of theater and moving picture.

TABLE III.—SEVENTY SENIORS WHO CHANGED THEIR OCCUPATIONAL CHOICE BETWEEN 1927 AND 1932; ALSO RANKINGS OF INTEREST SCORES FOR OCCUPATION GIVEN, WHEREVER AVAILABLE

| No. | Occupational choice in 1927 | Median ranking of occupational interest score in | | | Occupational choice in 1932 | Median ranking of occupational interest score in | | |
|---|---|---|---|---|---|---|---|---|
| | | 1927 | 1932 | Diff. | | 1927 | 1932 | Diff. |
| 1 | Engineer | 5.0 | 13.0 | −8.0 | Chemist | 7.0 | 13.0 | −6.0 |
| 1 | Engineer | 1.5 | 2.5 | −1.0 | Farmer | 1.5 | 1.0 | .5 |
| 1 | Engineer | 3.5 | 4.5 | −1.0 | Physicist | 15.5 | 1.5 | 14.0 |
| 1 | Engineer | 7.0 | 3.0 | 4.0 | Teacher | 4.0 | 9.0 | −5.0 |
| 4 | Engineer | 1.0 | 3.0 | −2.0 | Business—general | | | |
| 5 | Engineer | 3.5 | 4.0 | −0.5 | Sales engineering | | | |
| 1 | Engineer | 4.5 | 3.0 | 1.5 | Patent expert | | | |
| 2 | Engineer | 1.5 | 7.5 | −6.0 | Don't know | | | |
| 1 | Engineer | 6.5 | 9.0 | −2.5 | Production manager | | | |
| 1 | Chemist | 10.0 | 12.0 | −2.0 | Real estate | 1.0 | 2.0 | −1.0 |
| 2 | Physician | 8.5 | 7.5 | 1.0 | Chemist | 9.5 | 6.8 | 3.7 |
| 1 | Physician | 1.0 | 3.0 | −2.0 | Professor of pathology | | | |
| 1 | Psychiatrist | .... | .... | ..... | Psychologist | 5.0 | 2.5 | 2.5 |
| 1 | Geologist | .... | .... | ..... | Engineer (petroleum) | 8.0 | 2.0 | 6.0 |
| 1 | Geologist? | .... | .... | ..... | Office work | 16.5 | 18.5 | −2.0 |
| 1 | Petroleum industry | .... | .... | ..... | Real estate | 14.5 | 6.0 | 8.5 |
| 1 | Lawyer | 1.0 | 2.5 | −1.5 | Athletic coach | | | |
| 1 | Lawyer | 1.0 | 2.5 | −1.5 | Bank teller? | | | |
| 1 | Lawyer | 1.5 | 2.5 | −1.0 | Artist | 13.5 | 9.0 | 4.5 |
| 1 | Lawyer | 6.0 | 4.0 | 2.0 | Farmer | 10.0 | 4.0 | 6.0 |
| 1 | Lawyer | 2.0 | 1.5 | 0.5 | Teacher | 11.5 | 4.5 | 7.0 |
| 1 | Writer | 1.0 | 2.0 | −1.0 | Banking | | | |
| 1 | Writer | 1.0 | 1.0 | 0 | Purchasing agent | 18.0 | 12.5 | 5.5 |
| 1 | Advertiser | 12.0 | 6.5 | 5.5 | Finance | | | |
| 1 | Life insurance | 4.0 | 4.0 | 0 | Lawyer | 4.5 | 1.0 | 3.5 |
| 1 | Insurance adjuster | .... | .... | ..... | Lawyer | 2.5 | 5.0 | −2.5 |
| 1 | School man? | 6.0 | 3.0 | 3.0 | Lawyer | 1.5 | 1.5 | 0 |
| 1 | School man | 1.5 | 2.5 | −1.0 | Forest service | | | |
| 1 | School man | 3.5 | 6.0 | −2.5 | Bond salesman | | | |
| 1 | School man | 10.5 | 9.5 | 1.0 | Mathematician | 6.0 | 9.5 | −3.5 |
| 1 | School man | 10.5 | 1.5 | 9.0 | Professor of history | | | |
| 1 | Business—general | .... | .... | ..... | Office work | 8.0 | 4.0 | 4.0 |
| 1 | Business—general | .... | .... | ..... | Advertiser | 6.5 | 1.0 | 5.5 |
| 1 | Business—general | .... | .... | ..... | Certified public accountant | 9.5 | 12.0 | −2.5 |
| 1 | Business—general | .... | .... | ..... | Lawyer | 1.5 | 1.0 | 0.5 |
| 1 | Finance | .... | .... | ..... | Lawyer | 4.0 | 4.0 | 0 |
| 1 | Finance | .... | .... | ..... | Office work | 4.5 | 6.5 | −2.0 |
| 1 | Foreign trade | .... | .... | ..... | Physician | 6.5 | 2.0 | 4.5 |
| 1 | Foreign trade | .... | .... | ..... | Writer | 1.0 | 1.0 | 0 |
| 1 | Retailing | .... | .... | ..... | Writer | 1.0 | 1.0 | 0 |
| 1 | Diplomatic Service | .... | .... | ..... | Office work | 2.0 | 7.0 | −5.0 |
| 1 | Economist | .... | .... | ..... | Lawyer | 4.0 | 1.0 | 3.0 |
| 19 | Miscellaneous[1] | .... | .... | ..... | Miscellaneous | | | |
| | Median, thirty-five cases | 2.2 | 3.4 | −1.2 | Median, thirty cases | 6.0 | 4.0 | 2.0 |
| | Quartile, distribution | 2.5 | 2.0 | ..... | Quartile, distribution | 3.6 | 3.7 | |

[1] Miscellaneous group comprises: two geologists changing to contractor and "don't know"; three mineralogists changing to salesman, contractor and sales engineer; one petroleum industry to "don't know"; eight business in general to four salesmen and one each of the following: Contractor, finance, retailing and professor of economics; two in finance to retailing and salesman (oil); two foreign trade to "don't know"; and one insurance to salesman (not of insurance).

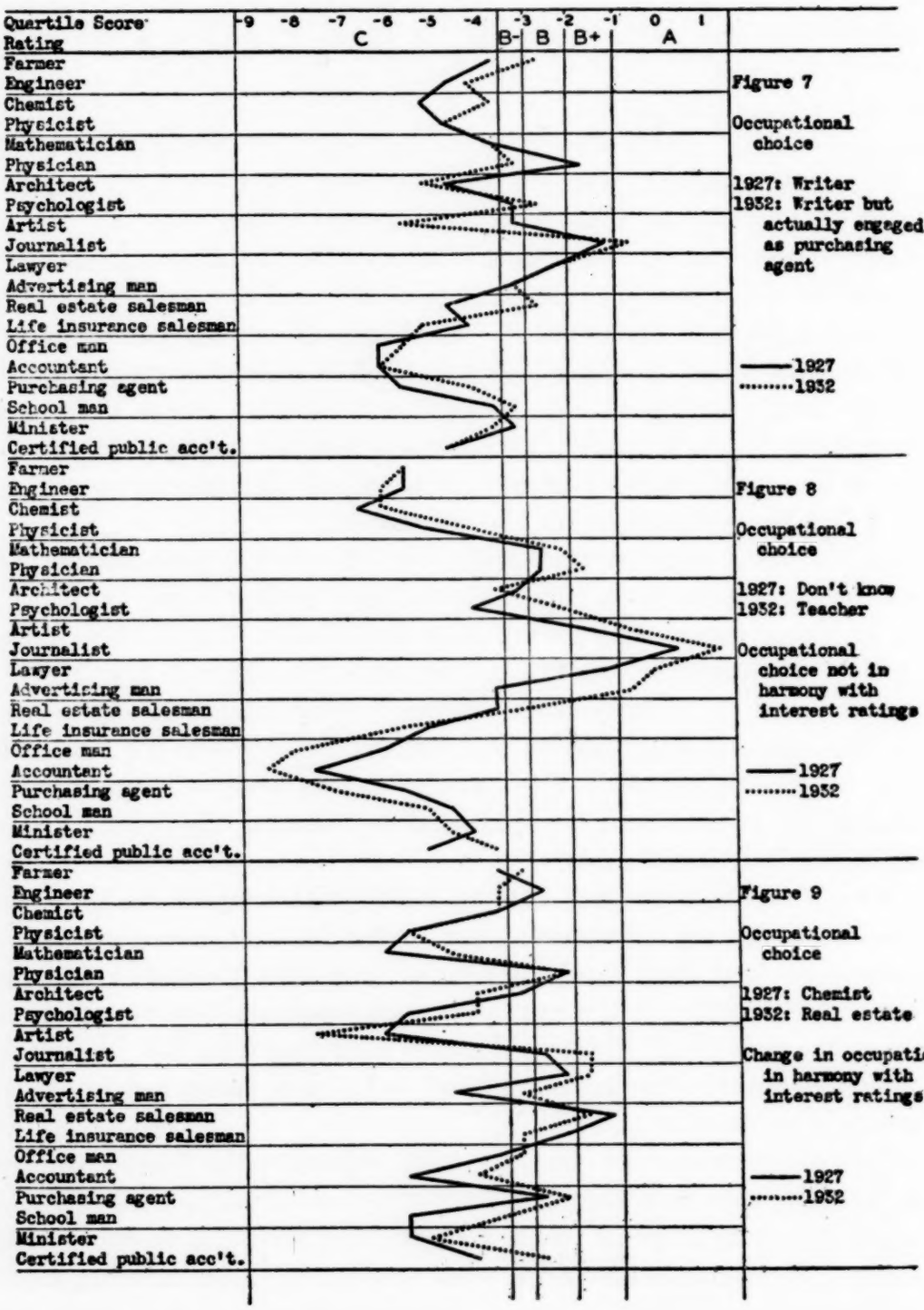
In 1932 these ratings were raised slightly, giving an *A* rating in engineering. The interest ratings for the average young lawyer are entirely different. He had a *B*+ rating in law, a *B* rating in journalism and selling real estate, and a *B*− rating in selling life insurance and purchasing. Here also these ratings are slightly higher in 1932.

The twenty-seven young engineers (Fig. 5) score two quartiles higher in engineering than does the average senior (Fig. 4). These two quartiles occur at a significant interval in the distribution representing the difference between ratings of *B*+ and *B*− in 1927 (ratings of *A* and *B* respectively in 1932). Similarly, the fifteen young lawyers score one and a half quartiles higher in the interests of lawyers than the average man (ratings of *B*+ in 1927 and 1932). These differences between the young engineers and lawyers and the entire group do not impress one as being very great, yet they are statistically significant (5.7 to 7.2 times the sigma of the difference).

All the data support very definitely the second proposition that men continuing in an occupation obtain a higher score in it than men entering some other occupation (see Table V and the various figures). For example, the fifteen lawyers score higher in law than the engineers, and the twenty-seven engineers score higher in engineering than the lawyers, with differences that are from 5.3 to 8.4 times the sigma of the difference. The proposition is true with respect to averages but it must not be construed as meaning that there are no exceptional individual cases.

Only three comparisons can be made with respect to the third proposition, namely, that men continuing in an occupation obtain higher scores in it than men who change from that occupation to some other. In the case of the engineers and teachers, the statement is true; in the case of the lawyers, the men changing from law to another occupation obtain a quartile score of −1.4 and those continuing in law have a score of −1.5.

The results are not so clean cut with regard to the fourth proposition, namely, that men changing from other occupations to occupation *X* score higher in *X* prior to the change than they did in other occupations. The proposition holds with respect to those entering medicine (see Table 5). In the case of those entering law the scores in law and journalism are tied for first place. In the case of those entering chemistry, the chemistry score is but 0.1 quartile lower than that for teaching. In the case of those entering teaching, the scores in law and journalism are distinctly higher than those in teaching, which


Quartile Score
Rating
-9 -8 -7 -6 -5 -4 -3 -2 -1 0 1
C
B- B B+ A
Farmer
Engineer
Chemist
Physicist
Mathematician
Physician
Architect
Psychologist
Artist
Journalist
Lawyer
Advertising man
Real estate salesman
Life insurance salesman
Office man
Accountant
Purchasing agent
School man
Minister
Certified public acc't.
Figure 7
Occupational choice
1927: Writer
1932: Writer but actually engaged as purchasing agent
——1927
·····1932
Figure 8
Occupational choice
1927: Don't know
1932: Teacher
Occupational choice not in harmony with interest ratings
——1927
·····1932
Figure 9
Occupational choice
1927: Chemist
1932: Real estate
Change in occupation in harmony with interest ratings
——1927
·····1932

ranks accordingly third, not first; and in the case of those entering engineering the scores in medicine and journalism are about 0.5 quartile higher than the score in engineering. Combining all five groups we find that men changing from other occupations to *X* average −2.18 in *X* and −2.10 in a second occupation (−2.71 in a third occupation). Consequently we must conclude that men changing from some other occupation to *X* score equally high in *X* and in a second occupation prior to the change than they did in any other occupation.

It is extremely unfortunate that among all these cases there are only fourteen men who have changed from one occupation to another, both of which can be scored on the interest test. A considerable amount of data of this sort would answer the question as to whether the interest test scores obtained while in college (1927) are more in harmony with occupational choice in 1927 or the occupation selected in 1932. As far as these fourteen cases go, the evidence is that the 1927 test is more in harmony with the first occupational choice than the second (median rankings of 4.5 and 6.5). After reading over the life histories of these men the writer expected to find the 1927 interest test scores much less in accord with the 1932 choice, because a considerable number are earning a living in occupations which they state they have entered because of economic conditions and not from desire. Figure 7 presents the case of a man, for example, who is a purchasing agent today, but who states very positively that his original ambition to be an author is still uppermost and that he is spending his spare time in writing. He rates highest in writing both in 1927 and 1932. If he continues as a purchasing agent he will be engaged in work not in harmony with his interests; if he finally achieves his present ambition, it will be a real vindication of the interest test scores. As it is, his rank of 1.0 in writing for both 1927 and 1932 and his very low ranks in purchasing (18.0 in 1927 and 12.5 in 1932) help explain why the interest test scores of 1927 do not agree quite so closely with occupational choice in 1932 as with that in 1927.

The foregoing data definitely sustain the first two propositions; the balance of proof is also in favor of the third proposition; in the case of the fourth proposition it is necessary to modify it to this extent, that men changing from other occupations to occupation *X* will score higher in *X* or one other occupation prior to the change than they did in other occupations.

### A SECOND METHOD OF HANDLING THE DATA

The second method of handling the data may be illustrated from Figs. 1 and 2. In the former figure the young lawyer has equally high scores in law and journalism in 1927 and these two occupations are ranked 1½. In 1932 law stands highest and is ranked one. These

TABLE IV.—FORTY-SEVEN SENIORS WHO REPORTED THEY DID NOT KNOW WHAT THEIR OCCUPATIONAL CHOICE WAS IN 1927 AND HAVE SELECTED ONE IN 1932; WITH RANKINGS OF INTEREST SCORES FOR OCCUPATION GIVEN WHEREVER AVAILABLE

| No. | Occupational choice in 1932 | Median ranking of occupational interest score in | | |
|---|---|---|---|---|
| | | 1927 | 1932 | Difference |
| 3 | Engineer | 8.0 | 5.0 | 3.0 |
| 1 | Chemist | 1.0 | 1.0 | 0 |
| 4 | Physician | 2.3 | 2.8 | − .5 |
| 1 | Architect | 4.5 | 4.5 | 0 |
| 2 | Farmer | 6.0 | 3.8 | 2.2 |
| 1 | Geologist | | | |
| 1 | Lawyer | 3.5 | 2.5 | 1.0 |
| 1 | Writer | 1.0 | 1.5 | − .5 |
| 3 | Advertiser | 8.5 | 10.5 | −2.0 |
| 3 | School man | 9.0 | 2.5 | 6.5 |
| 5 | Office man | 3.5 | 2.5 | 1.0 |
| 1 | Accountant | 2.5 | 1.0 | 1.5 |
| 21 | Miscellaneous[1] | | | |
| | Median, 26 cases | 4.4 | 2.7 | 1.7 |
| | Quartile, distribution | 3.3 | 2.5 | |

[1] Miscellaneous group comprises: Six business—general, four finance, one retailing, four salesmen, one contractor, four college professors, one dramatics.

two ranks of 1½ and one are used here to represent the relationship between the occupational choice of this man and his two interest test scores. Such a procedure has real significance since if he were accepting the interest test scores at face value he should enter the occupation on which he scores highest. In the case of the young lawyer in Fig. 2, law ranks three in 1927 and six in 1932. As each blank has been scored for twenty occupations the ranks range between one and twenty.

The median of all such ranks for all the men where it is possible to score their blanks for the occupation chosen, one hundred seventeen men all told, is 1.89 on the 1927 test and 2.07 on the 1932 test (Table VI). As the difference between these two medians is insignifi-

TABLE V.—VOCATIONAL INTEREST SCORES IN 1927 OF: (1) MEN CONTINUING IN THE SAME OCCUPATION FROM 1927 TO 1932, (2) MEN CHANGING FROM GIVEN OCCUPATION TO ANOTHER ($X$), AND (3) MEN CHANGING FROM ANOTHER OCCUPATION ($X$) TO THE GIVEN OCCUPATION

| No. | Occupational choice in | | Quartile scores in | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1927 | 1932 | Engineer | Chemist | Physician | Journalist | Lawyer | Real estate salesman | Minister | Teacher | Purchasing agent |
| 27 | Engineer | Engineer | −1.3 | −1.9 | −3.9 | −3.6 | −4.0 | −5.1 | −6.0 | −4.2 | −2.7 |
| 19 | Engineer | $X$ | −2.1 | −2.7 | −4.3 | −3.5 | −3.8 | −4.8 | −5.2 | −3.6 | −2.8 |
| 3 | $X$ | Engineer | −2.8 | −3.1 | −2.3 | −2.6 | −3.1 | −4.3 | −4.8 | −3.8 | −3.3 |
| 12 | Chemist | Chemist | − .5 | .2 | −2.3 | −3.2 | −3.6 | −6.1 | −5.5 | −3.7 | −3.2 |
| 6 | $X$ | Chemist | −3.4 | −2.6 | −3.8 | −3.9 | −3.6 | −5.2 | −4.3 | −2.5 | −3.6 |
| 17 | Physician | Physician | −3.6 | −2.7 | −1.7 | −2.6 | −2.7 | −4.8 | −4.5 | −3.4 | −4.5 |
| 6 | $X$ | Physician | −2.2 | −2.2 | −1.2 | −2.6 | −2.6 | −4.8 | −5.3 | −3.8 | −3.7 |
| 15 | Lawyer | Lawyer | −4.3 | −6.7 | −5.4 | −2.1 | −1.5 | −2.8 | −5.4 | −4.1 | −3.8 |
| 8 | Lawyer | $X$ | −5.4 | −6.5 | −5.1 | −1.9 | −1.4 | −2.9 | −5.2 | −3.7 | −3.8 |
| 7 | $X$ | Lawyer | −5.3 | −5.4 | −5.2 | −1.8 | −1.8 | −3.1 | −5.4 | −4.2 | −4.0 |
| 21 | Teacher | Teacher | −4.4 | −5.3 | −4.1 | −2.5 | −2.4 | −4.8 | −2.9 | −1.0 | −4.4 |
| 7 | Teacher | $X$ | −3.6 | −3.7 | −4.1 | −2.9 | −3.3 | −4.8 | −4.4 | −2.0 | −3.7 |
| 5 | $X$ | Teacher | −5.6 | −6.9 | −5.0 | −1.8 | −2.2 | −4.6 | −4.6 | −3.0 | −4.4 |

cant we can conclude that occupational choice in 1927 agrees approximately equally with an interest test administered at the time that choice was made or five years later. The results indicate also that one-fourth of these men have selected an occupation which is ranked one ($Q_1$ of 1.09 in 1927 and 1.21 in 1932), that one-half have selected an occupation which is ranked two or better (medians of 1.89 and 2.07 respectively) and three-fourths have selected an occupation which is ranked four or better ($Q_3$ of 3.73 and 4.07 respectively). This means

very close agreement between occupational choice in 1927 and interest test scores at that time or five years later. The relationship is better than this for the eighty-two men who have not changed their occupational choice during this period and it is somewhat poorer for the thirty-five men who have changed their occupation.

TABLE VI.—RANK OF 1927 OCCUPATIONAL CHOICE IN TERMS OF INTEREST TEST SCORES

| | No. | 1927 interest test | | | 1932 interest test | | | Difference between medians | Critical ratio[1] |
|---|---|---|---|---|---|---|---|---|---|
| | | 1st *Q* | Med. | 3rd *Q* | 1st *Q* | Med. | 3rd *Q* | | |
| No change in occupational choice..... | 82 | 1.08 | 1.78 | 3.05 | 1.10 | 1.65 | 3.38 | .13 | 0.7 |
| Change in occupational choice..... | 35 | 1.11 | 2.17 | 6.06 | 2.33 | 3.40 | 6.38 | −1.23 | 1.8 |
| Summary..... | 117 | 1.09 | 1.89 | 3.73 | 1.21 | 2.07 | 4.07 | − .18 | 0.8 |

[1] Critical ratio is $\frac{\text{difference between medians}}{\text{quartile of difference}}$. Ratios above 4.0 are accepted as statistically significant.

TABLE VII.—RANK OF 1932 OCCUPATION ENGAGED IN IN TERMS OF INTEREST TEST SCORES

| | No. | 1927 interest test | | | 1932 interest test | | | Difference between medians | Critical ratio |
|---|---|---|---|---|---|---|---|---|---|
| | | 1st *Q* | Med. | 3rd *Q* | 1st *Q* | Med. | 3rd *Q* | | |
| No change in occupational choice..... | 82 | 1.08 | 1.78 | 3.05 | 1.10 | 1.65 | 3.38 | .13 | 0.7 |
| Change in occupational choice..... | 30 | 2.63 | 6.00 | 9.88 | 1.44 | 4.00 | 8.81 | 2.00 | 1.7 |
| Change from "don't know" in 1927 to definite choice in 1932..... | 26 | 1.81 | 4.38 | 8.44 | 1.58 | 2.73 | 5.63 | 1.65 | 1.6 |
| Summary of last two groups.... | 56 | 2.08 | 5.00 | 8.92 | 1.54 | 3.00 | 6.75 | 2.00 | 2.8 |
| Summary of all three groups... | 138 | 1.19 | 2.36 | 5.10 | 1.19 | 1.97 | 4.83 | .39 | 1.4 |

The preceding paragraph establishes the fact that what a college senior claims he wants to do is shown very accurately in his interest scores at that time and also five years later. Consider now the relationship that exists between the occupation he is actually engaged in five years after graduation and his interest scores at that time (1932) and five years earlier. This is shown in Table VII. For the entire group of one hundred thirty-eight men the occupation engaged in ranks 2.36 in 1927 and 1.97 in 1932 (critical ratio of the difference is only 1.4). Here again there is good agreement between either interest test and occupation engaged in in 1932, but not quite so good an agreement as was found in connection with occupational choice in 1927. The difference is largely to be accounted for on the ground that some men have entered an occupation not from choice but from economic necessity.

The data for all the one hundred thirty-eight men are influenced considerably by the eighty-two men who have not changed their occupation at all. If the other two sub-groups are alone considered it appears that the 1932 interest test agrees better with 1932 choice than does the 1927 interest test (medians, respectively, of 5.0 and 3.0), but the difference of 2.00 has a critical ratio in terms of quartiles of only 2.8, not large enough to be deemed statistically significant. Upon the basis of these figures it appears that when a man changes his occupation he will obtain a slightly poorer interest score in the first occupational choice after the change than before (median rankings, respectively, of 2.17 in 1927 and 3.40 in 1932). And similarly apparent is the converse tendency to obtain a slightly better interest score in the second occupational choice after it has been selected than before (median rankings, respectively, of 5.00 in 1927 and 3.00 in 1932).

Every new experience enriches our understanding of terms and it is only to be expected that some items on the blank will take on new meanings after entering an occupation and so be checked differently. Certain other changes in marking the blank can be explained by the tendency to mark them as one ought to rather than in terms of one's real interest. Thus men applying for a job to sell insurance nearly always mark that item as liked, although many of them are rated *C* in that interest. The amazing thing is that the total interest score is so little affected by such reactions. For example, the median ranking in 1932 of 1.65 is only .15 higher than that in 1927 for the eighty-two men who did not change their occupational choice. Five years actual experience seems to have had little effect in this respect.

## INDIVIDUAL CASES

General conclusions must be based on averages, as has been done in the preceding sections. In interpreting such conclusions it is always important to consider the variations that are to be found among the data. Several such examples are accordingly reported.

The record given in Fig. 8 illustrates how greatly occupation engaged in can differ from interest ratings. This young man, a major in English in college, reported in 1927 that he did not know what he was going to do. In 1932 we find him teaching in high school. This occupation ranked 11th in 1927 and 12½ in 1932. Journalism is clearly first in both tests. In 1927 his day-dream was "music (voice—preferably as performer)[1]—the one dream has been, and still is, opera." In 1928 he states he is teaching English, Spanish and history to earn money to go to Germany to study music. In 1929 he reports he is planning to prepare himself for college teaching of English; that he had wanted to sing but the cost far surpassed his ability to pay for it. In 1932 he hopes to obtain a Ph.D. in English preparatory to a university appointment. He reports he is considered a success but detests most of what he has to do. His low interest test scores in teaching and his high scores in music and journalism are in line with his desires.

Figure 9 illustrates the case of a young man who has already changed his occupation in harmony with his interest records. His first choice, chemistry, ranked ten in 1927. In 1929 he tried medicine (rank 2½) attending the Harvard Medical School that year. Then he turned to the real estate business, in which he is now engaged. This latter occupation ranked first in interest ratings in 1927 and is tied for first place with journalism and law in 1932.

## CONCLUSION

1. One-half of college seniors change their occupation within five years after graduation. Some of these changes represent no loss, for the experience gained is worth the time spent in terms of future advancement. But in many cases the change means a genuine loss of time and earning power, as can be readily appreciated by considering the examples in Table III.

---

[1] He had a rating of *B*+ (almost an *A*) as musician and *B* as artist in 1927.

2. There is close agreement between the occupational choice of college seniors (when they have a choice) and interest test scores at that time and five years later. This is true on the average but does not hold true of every individual case.

3. There is some change in interest test scores between the senior year and five years later. The latter are slightly more in agreement with occupation engaged in five years after graduation than the former but the difference is not statistically significant.

4. Men continuing in an occupation obtain a higher score in it on the interest test than in any other occupation.

5. Men continuing in an occupation obtain a higher score in it than men entering some other occupation.

6. Men continuing in an occupation obtain higher scores in it than men who change from that occupation to some other.

7. Men changing from some other occupation to occupation *X* five years later score higher while seniors in *X* or one other occupation than they did in eighteen other occupations.

8. Men successfully employed in an occupation are differentiated from other occupational groups by the vocational interest test. Young men who obtain interest scores indicative of membership in an occupational group are more likely to enter the occupation than any other.

# TYPES OF ASSUMPTIONS IN EDUCATIONAL RESEARCH[1]

DOUGLAS E. SCATES

Director of School Research, Cincinnati

All science is based upon a hierarchy of assumptions, which cover a wide range in kind and degree. At one extreme are the readily tested assumptions made by a worker in a minor phase of his research, and at the other are the basic postulates of knowledge which are neither indubitably true nor demonstrably false. These assumptions are prevalent in science because they are common in the research which produces science. Some of them are unavoidable, some are contemporaneously necessary, some are unwarranted. It is the purpose of this paper to examine briefly a number of the assumptions which commonly enter into educational research, and to offer a classification of them. The paper is practically limited to research which utilizes statistical methods and seeks to study associated variation. Such a treatment, without pretending to be exhaustive in a detailed way, should be of value in focussing attention on assumptions generally, in suggesting certain assumptions which may tend to be overlooked, and in developing a broader perspective from which to view individual assumptions that are noted.

The assumptions which enter into educational research will be treated under four main types, with a number of classes under each type. These types will be designated as (1) technical, (2) methodological, (3) social, and (4) philosophical, and will be discussed in this order, on the basis of their increasing generality. Illustrations of each type will be given which should serve to suggest other assumptions not enumerated.

An assumption is a mental datum which is not fully established, but which is used as a basis for continuing the thought or study. The term "assumption" covers all facts, principles, or other concepts, the truth of which is taken for granted for particular purposes without

[1] Read before the Ohio Conference of Statisticians, Cincinnati, November 5, 1932. Two other papers read at this conference, and dealing with assumptions in particular fields of research, have been published as follows: Tyler, Ralph W.: "Assumptions Involved in Achievement-Test Construction." *Educational Research Bulletin*, Vol. XII, February 8, 1933, pp. 29–36. Hendrickson, Gordon: "Some Assumptions Involved in Personality Measurement." *The Journal of Experimental Education*, Vol. II, March, 1934, pp. 243–249.

insistance upon specific proof. Conditions which are assumed may or may not be true; in the field of research, however, there is certainly the expectancy that one believes them to be true before he proceeds to act upon them. There is further the presumption that any assumption which has come to the attention of the worker has been examined as carefully as the circumstances permit, and that if the assumption can not be verified, it will be stated in connection with, and as a part of, the interpretations and conclusions of the study. It is probable that no research worker can be aware of all of the assumptions on which he depends in the process of his work, but the careful worker will be constantly on the lookout to discover them.

The technical assumptions, which comprise the first type, are fairly common and easy to recognize. They occur in the derivation and use of the more formal procedures in every field of study. They are made in the course of research to enable the worker to proceed toward his goal by utilizing data or relationships which are advantageous to him even though they are not known to be correct. If these assumed data or relationships could not be substituted for those which were known to be true, either the progress of the work would be substantially encumbered, or it might be blocked altogether. These technical assumptions can probably be put into three classes: (1) assumptions of numerical equality for approximations which are never, or almost never, exactly true; (2) formal assumptions of condition, which frequently underlie standardized statements of relationship which are true only under certain limiting conditions; (3) informal assumptions of use which are implicit in the manner in which raw or derived data are used by the research worker.

Numerical approximations in statistical work are rather common. They may occur in the data, in the calculations, or in the formulas used. For example, the data with which one works are ordinarily approximations owing to the limits of accuracy in the measuring instruments and in the observations, and to the crudity of units used. Further, data are usually grouped into class intervals and their approximate average values used for calculations. Final results are rounded, or the calculations not carried out. Successive approximation methods are used to evaluate constants where direct attack is difficult. In the case of formulas, those used to calculate probable errors are almost invariably approximations to longer, more exact, expressions. Blakeman's formula may be cited as a familiar illustration in which a coefficient of several terms is assumed to be equal to unity and is

disregarded, although it is only exactly equal to unity when eta and *r* are equal. Many of these rather obvious approximations are probably something of a shock to the student just entering the field of research, with a love for exactness, but further experience and a more mature perspective make clear that they are not likely to be serious in comparison with the less conspicuous but less limited uncertainties that lie in the original data and in the analysis and interpretation of these data.

The second group of technical assumptions—those which require that certain specific conditions be met in order for the relationship to be true—are also numerous in statistics. For example, the Pearson product-moment correlation method assumes that the two distributions are linear when plotted on each other; if they are not, the correlation coefficient underexpresses the degree of functionality present. The formula for rank correlation holds only under the condition that there are no ties in rank in either of the series. Many of the formulas for probable error are valid only when the distribution of the constant under the influence of sampling is normal. The short formula for the probable error of a difference assumes that the values are uncorrelated. The Spearman-Brown formula requires that the material be "similar" or homogeneous throughout the range to which it is applied. The chi-square test assumes that the deviations will not lie systematically on one side, that the grouping will not be too fine, that the population will be large, and other things. These and similar assumptions are regarded as formal because they are necessary parts of the relationships which are expressed in formulas. Unfortunately they do not always appear with the final statements of formulas and are therefore sometimes lost sight of.

These formal assumptions of condition are typical of those which workers engaged in the production of new tools of statistical analysis are called upon to make. They are created by the workers in pure research as the need for them arises. They are by-products of rigorous reasoning based upon conditions shaped to suit the occasion. A worker may have to exercise considerable ingenuity to discover assumptions which will enable him to proceed; but once they are found, he has merely to note them and go on. Those who make use of these research tools must accept these assumptions that have previously been made and endeavor to meet them, for every assumption which enters into the derivation of a formula or other standardized relationship must enter into the use of that relationship, and the final conclu-

sions must be dependent upon the extent to which such assumptions are known to be met. Translating these abstract assumptions into practical working terms often proves difficult. If the use of the Spearman-Brown formula requires that material must be homogeneous, the user must be able to interpret that requirement; in what ways must it be homogeneous? Obviously it must differ in some respects if it is not to be identical. After large abstract assumptions have been interpreted, the practical worker must know the indicia of various characteristics of his data which will indicate whether the required conditions are met. In some cases there are formal checks; in others an inspection is satisfactory, while in still other cases the worker must depend on his knowledge of the underlying factor as a basis for judgment.

Somewhat similar to the second group of technical assumptions is a third group, called assumptions of use, which are regarded as informal because they do not enter into the derivation of a formula but rather grow out of the particular use to which calculated values or analytical procedures are put. While they are, like the second group, assumptions of condition, they are not normally made by the producer of research tools but rather by the user of these analytical tools. They are created in the process of solving practical problems. A simple illustration is the assumption underlying the *a priori* judgment of a worker that it is immaterial whether he calculates the mean or the median to index the central tendency of a given series; he thereby assumes that the distribution is symmetrical within the limits of error permitted by his purpose. The assumption is obviously not inherent in either constant (statistic), but grows out of the relationship in which the two constants are placed by the judgment of the worker.

It should be clear that there are important assumptions connected with the use of all statistical constants and analytical procedures. The proper use of statistical constants calls for an understanding which is broader than a knowledge of the assumptions which enter into the calculation of those constants. For example, the use of averages for direct comparison of distributions implies that other moments of the distributions are reasonably comparable. The use of the median in preference to the mean rests on the assumption that one is more interested in a value which reflects rank than in one which reflects moments. The use of the standard deviation as a unit for equalizing the apparent weighting of several series which are to be combined rests upon the assumption that higher moments of the various series

are approximately equal, and that the intercorrelations are approximately equal. Correlation may be used to show the presence or absence of relation between two traits with only formal assumptions involved; but when it is used to express one variable as a function of another, it rests upon the assumption that other conditioning (independent) factors are of secondary, or negligible, importance. Furthermore, the use of all of these constants, excepting the averages, without correction for attenuation, makes the assumption that errors in measurement are not of great consequence. This assumption is involved for example when the observed value of a standard deviation is used in calculating a probable error. Again, the use of a probable error for a series constant (statistic) rests on the assumption that one has a random sample of a particular universe, and wishes to estimate limits having a certain probability, for corresponding constants of samples similarly drawn from the same universe.

These assumptions of condition, constituting the second and third groups of technical assumptions, are the source of considerable uncertainty. While the error arising from them is likely to be greater than the error in the first group (numerical approximations), its magnitude is frequently difficult to estimate. In some cases of practical work it is possible only to set probable extreme limits for the error, for the assumption must be made largely or wholly in the dark. In the calculation of biserial $r$, for example, one assumes that a dichotomy represents two portions of the area of a normal distribution, and in some cases one might know nothing at all about the real shape of the distribution. (One can, of course, normally avoid the totally erroneous application of the assumption to a dichotomy of *kinds* of cases which overlap each other on the scale, such as boys and girls on intelligence or on most any other trait.) The same difficulty may occur when one indexes a qualitative series by assigning mean values of portions of a normal curve; the assumption of normality of distribution, though frequently as good as any other assumption, may be wholly gratuitous.

The problems of correlation are particularly intriguing. If the magnitude of various correlations is to be compared, the calculations must be made upon corresponding ranges of the several variables concerned. In well explored fields this may be possible, though it frequently lies beyond the means of an individual research worker. In new fields, where the variety of characteristics has not been explored and homogeneous populations have not been isolated, it is practically impossible, since one does not know what ranges to expect. The

condition is generally difficult to meet satisfactorally when dealing with series in which time between observations is a significant factor, since fundamental conditions may change greatly over a period of time, and yet they may change so gradually that it is difficult to delimit periods that may be regarded as homogeneous. For example, one may correlate the number of school buildings in the United States with the total number of pupil-days of schooling per year. For many years both increased; then small buildings began being replaced by large ones and now the number of buildings is decreasing while the aggregate pupil-days of schooling is still increasing. How would one determine the period within which the range of these variables was equal to the range of some totally unrelated variables with which he wished to compare the degree of correlation? Many of the assumptions involved in the use of correlation, of one kind or another, must fall in the class of those which can scarcely be definitely checked.

Some assumptions of condition can be studied for general tendency without yielding to direct attack in a particular situation. Response error formulas are developed on the assumption that errors of measurement are not correlated with each other or with the values of the variable. We may be able, through research directed to that end, to ascertain whether, for a particular kind of measurement, this is likely to be so, and in a particular instance we may employ the assumption, relying upon the general probability of its being met. Again, when the average response error of a group is attributed to an individual, or in fact when the variability of any individual measure is inferred from the variability of the group, the assumption is made that the variability of the individual under successive sampling procedures is the same as the variability of the group in a given set of observations. In biological fields, this is tantamount to saying that the factor that causes variability in the group at a given time and place will cause variability in an individual from time to time or from place to place. However absurd this may be, the assumption continues to be tacitly made. In somewhat the same class is the interpretation of the product-moment correlation in terms of per cent of common elements, usually requiring assumptions of independence and a probability of one-half for the elemental factors underlying the variables. Research can establish the general probability of the conditions which are assumed to be true in these and other similar cases.

A further difficulty in determining the amount of error arising from assumptions of condition is to be found in the fact that we do

not know in many cases the seriousness of a given departure from the assumed conditions. That is, granting the availability of checks by which one may determine how well his assumptions are met, he still may not know the extent to which straightforward conclusions are invalidated when the checks reveal a negative result. Blakeman's test will reveal whether departure from linearity in a frequency table is beyond the probable limits of chance, but it tells nothing about the effect of this departure upon the value of *r*. Noting that ties in rank affect the value of rho, one is still uncertain as to the point at which he should abandon the technique. Variations in the specified conditions may or may not alter the stated function rapidly; if the conditioning functions and the stated function are closely correlated, then a marked departure from the truth may be expected when the predicated conditions are not met. If, on the other hand, the two sets of functions are not closely related, particularly in the area in which calculations are usually carried on; then a considerable degree of tolerance in the necessary conditions may be permitted without incurring large error. The extent to which it is possible to depart from assumed conditions without exceeding a specified amount of error appears to be a field for profitable study.

To many workers in research, technical assumptions are of outstanding concern. The technical aspects of research are not however ordinarily of the most importance in determining the value of research. The processes which exert the greatest influence on the essential quality of research are those which are here classed as methodological. In this category we place those larger acts of judgment by which research is guided, and those insights which are so necessary if research is to be vital. On the level of methodology we contemplate the efficiency of the general plan of attack on a problem, the circumspection with which the data are secured, the sagacity with which the analysis is prosecuted, the accuracy with which the interpretation and conclusions are drawn. For our immediate purpose, we are concerned with the assumptions that are likely to be involved in these processes.

The detailed assumptions underlying the methodological phases of research are so numerous, and vary so much both with the kind of problem and the field of work, that only the more fundamental assumptions can be dealt with. Perhaps the most basic assumption, for research which deals with concomitant variation, is that the data which one secures contain, in extractable form, evidence of the respective amounts and kinds of influence that are exerted by factors which

are crucial and adequate to the solution of the stated problem. This condition is a necessary prerequisite to conclusive research, and yet it can scarcely be established beyond a peradventure. One can often examine his data for reasonableness and consistency, but after all analysis is primarily forward looking; it is designed to bring to light characteristics of the data which should answer the principal question of the study if these characteristics actually represent the influences which they are presumed to. The status of the fundamental assumption must rest upon a number of supporting assumptions, and ordinarily these must await substantiation by further research in this same field.

Of large importance in determining the significance of the basic data is the worker's concept of the problem. According to his insight (and perhaps fortune) in perceiving in advance the nature of the contributing factors and their relationships to the criterion variable will he be enabled to formulate a working hypothesis that proves efficient in the reduction of the main problem. Obviously all of his work that is carried through on this plan hinges upon the appropriateness of this primary working assumption. This is not to suggest that the plan will not be altered in the light of increasing familiarity with the field as the work progresses, but these adaptations are more in the nature of modifications than fundamental alterations in the attack. If one is directing his efforts toward the discovery of a certain relationship which does not exist, or which does not exist in the form in which he is seeking it, or between the particular variables with which he is working, he must broaden the scope of his attack, or start anew to gather observations, or be disappointed. On the other hand, the worker who tries to avoid such assumptions in his work by carrying on his research without hazarding an advance formulation of an hypothesis which he thinks may hold the solution to the problem, is likely to be going through rather mechanical motions that will not penetrate deeply.

When one selects (includes) certain factors for study, he assumes that they are at least likely to exert a significant influence on the phenomenon constituting the problem; that they are reasonably stable and can be consistently identified; that some function can be found which will render their variation objective without too much loss of fidelity; that the factors can be isolated from other factors which lie outside the limits of the study; that their variation can be measured with sufficient refinement to register minor but significant changes;

that the types of measurement currently available (and employed) will reveal the many different kinds of influence which these factors may exert; that the interrelations of the factors can be adequately dealt with; and so on. The consumer of research results is expected to assume that, insofar as appropriate, all of these things *have* been done.

In carrying forward a study the source and distribution of observations are of basic importance. The source of the data must be so located (or prepared and controlled) as to contain the factors which are sought, and it must be so extensive as to permit an effective range of operation for these factors, without being so broad as to introduce unnecessary extraneous factors, with perhaps the result that normal effects are offset and "average out." Within such more or less indistinct limits one must secure a distribution of observations of the several variables which is representative of the conditions in this source-field. To be representative, the sample must be properly proportional in its several significant aspects, and usually the nearest practical approach to this is to be secured through an attempt to obtain a random sample. When one speaks of the breadth of the source-field, he must have in mind extensiveness with respect to time, distance, and force, and the various complex conditions within the field of the study which are associated with these fundamental variables. The requirement for the proper distribution of observations is often complicated by the fact that the source-field for the results is not coterminous with the source-field for the conditioning factors, since many a consequence occurs at a point far removed in time or space from the effective conditions. This situation makes many psychological and social studies well nigh impossible because of the difficulty of controlling intervening conditions, keeping track of cases, locating cases with proper antecedents, etc.

Without continuing the review into unreasonable tedium, it should be obvious that one can, with a little thought, extend to considerable length the list of conditions required for satisfactory research, and that in actual practice each of these conditions normally resolves itself into one or more assumptions. In any complex study it is practically impossible to ascertain definitely whether the majority of the important conditions have actually been met. Their status must depend upon the constant watchfulness and critical judgment of the worker, based upon an intimate knowledge of the field being studied, and reinforced with a disciplined acumen and a scientific rigor. Under favorable

conditions, the assumptions become high probabilities; but they must be recognized as conditioning elements that cannot be entirely removed from research.

The two classes of assumptions which have been discussed, namely the technical and the methodological, perhaps constitute the groups that normally come to mind when the term "assumptions" is mentioned. Important as they are, however, they do not include all of the assumptions that are involved in the propagation of research. Many problems connected with research lie entirely outside the field being studied. The worker himself is a complex variable, responding strongly to the social forces which play upon him. The work that he does is in large part the product of the society in which he finds himself. It is necessary therefore to consider certain characteristics of his society for they will determine in a large way the dynamics of the worker, his likelihood of making a significant contribution, and the general orientation of his undertakings.

When one undertakes a piece of research, on his own initiative, he does so because he feels that it is somehow worth while for him to do it. This feeling may grow out of a direct personal satisfaction in the work, but ordinarily it is in part dependent upon the expectation that the results of one's work will sooner or later be found of value by certain elements of his society. The assumption here is a double one. In the first place the individual must feel that he can bring to the task a set of abilities that will enable him to make a reasonably unique or at least novel contribution. In the second place he must feel that society desires an expanded science in the particular field of his work, or that his results will be kept alive until this social frame of mind does develop. While these are generally taken for granted by trained research workers of our country, we must not overlook the fact that neither of these conditions has been true at all times since the dawn of the historical era, and that there are still large portions of the world in which they do not obtain.

Most of the oriental civilizations have not yet thoroughly acquiesced in the notion that inductive science is desirable. We need not dwell on the large groups of uncivilized peoples. Even in our own country, we cannot take it for granted that all research proceeds with the assurance of acceptability on the basis of its quality. There are a number of significant areas in which research is positively unwelcome to the controlling groups of our people. Outstanding in this respect are religion, politics, home life, and other forms of relationship looked

upon as personal. There is also great potential opposition from certain vested interests, some being well organized, which would view with considerable alarm any significant work that was proposed to be done in their fields. Research which is undertaken in these areas must recognize the contingencies in the assumptions previously stated. The acceptability of a good deal of possible research must yet wait upon the power of basic social forces to overcome ignorance, apathy, opposition, and the inertia of generations of social custom.

Even the ability of the worker is contingent upon the scientific interests of society. The fine backgrounds and skills which many research workers in a great variety of fields are able to bring to their complex problems grow out of the environments in which they have been privileged to live, being contributed to by training in universities, the reading of professional literature, and direct experiences with the work of contemporaries in the field. The barren periods in the past history of scientific progress do not signify that only people of mean capacity were born; they are known to be periods in which society generally was not interested in the cultivation of intellectual curiosity and its contributions and they point to the dependence of individual ability upon a rich supporting environment. Not only is research starved and forced into the marginal time of a few scattered workers when society is disinterested or hostile to it, but if this condition continues for a considerable length of time the competence of individuals disappears, gradually to be recreated later out of a more favorable social background. When one undertakes a research of wide significance, therefore, he is making the assumption (and often naively) that the environment in which he has been cultivated has permitted him to develop abilities that will enable him to do a piece of work that will be regarded as constructive by other scientists, throughout a wide area involving perhaps a number of differing and more or less competitive cultures.

One of the very subtle ways in which the mores of one's civilization control his efforts in research is to couch his thinking in terms of the larger social practices to which he has become used. The problems which are investigated are typically those which constitute details in the general practices of one's people, rather than those which raise basic questions concerning these practices. In the field of education, for example, it has been common for workers to assume, among other things, that education should be compulsory from about five or six to about sixteen or eighteen years of age; that chronological age is a

satisfactory basis for determining the limit of compulsory attendance; that any education outside of these limits is largely gratuitous; that all children will study the same subjects in approximately the same manner; that the pupils will be graded, advanced from grade to grade, etc. We must grant that, if the results of research are to be immediately useful, most of it should fall within the bounds of current usage. It must be recognized on the other hand that all such work rests upon the assumption that the possibilities which remain untested are not of great significance. It is well for research to be adjusted to felt needs, but it is also the larger province of research to discover needs which are as yet unsensed, and to extend its field beyond the limited concern of those who think only of the present.

The large conditioning factors which have just been discussed are fundamental in the research that is produced in any society. They control the support for research, determine the market for its products, afford the background of scientific insights which make it possible, and exert a powerful influence on its character. When one undertakes research of any significance, therefore, he assumes a good many things of practical importance. He takes for granted that groups of people in strategic positions will or will not be interested in certain problems that he might work on; he takes for granted that he can make a worth while contribution to the existing knowledge in the line of his endeavor; and, along with many assumptions of a lesser character, he ordinarility takes for granted that research which is premised on the customs of his group is contributing to social welfare. These assumptions are of course ordinarily unverbalized, and tend to be omitted from the consideration of those who are inclined to think only in terms of the more narrow scientific aspects of their work. They are here regarded, however, as of sufficient importance to constitute by themselves a third distinct type of assumption underlying research, and are denoted as social assumptions.

There is a still broader and more fundamental type of assumption which underlies all research and all science. This fourth type comprises a group of basic postulates which are classed as philosophical. These have to do with the assumptions underlying perception and cognition. For centuries logicians have endeavored to establish the characteristics of a science which would be perfect in its objectivity and completeness. The proposals which have been made have ranged from the one extreme of intellectualism, in which all knowledge is regarded as originating in pure reason, to the other extreme of

objective realism, in which all knowledge is regarded as having objective existence independent of the human mind. As a consequence society has gone through alternate periods of intense formalism in abstract reasoning, and equally intense devotion to sense-realism, with the result that we have today many who emphasize thought processes almost to the exclusion of observation, and many others who insist on objectivity almost to the exclusion of thinking.

It should be apparent, however, that an ultimate science cannot be established solely through the exercise of pure reason, nor again solely through a mechanical and presumably infallible type of observation. If we should depend entirely upon the mental processes of induction and deduction, both the means and the results must forever remain postulates, for we have provided no certain methods of verifying them, and it cannot be assumed that the human mind is incapable of error. If we should depend entirely upon objectively collected data, without mental processes, we should have the merest fragments of a science, for the interrelations of nature are apparent only as they are reconstructed within the human mind as hypotheses and theories which are not inconsistent with observation. It is, moreover, highly irrational to contemplate observation apart from the influence of the human mind, for any process of perception is dependent upon the observer as well as upon the external fact. The conception of a science which is entirely free from the possibility of human bias becomes therefore an absurdity. We are forced to depend for our science upon the interplay of mind and object, through the medium of perception.

This mind-object relationship gives rise to a number of assumptions which are of considerable interest. First among these is the assumption that science in a particular field is possible. In most instances this is taken as synonymous with the assumption that a science of the usual kind is possible, which is tantamount to assuming that there are laws covering the relationships between the variables, that these laws can be discovered and described, and that such a system of descriptions will embrace the most worth while aspects of the field.

Because of the success that scientists have had in certain fields they are inclined to assume that there is some orderly organization underlying phenomena in all fields. It must not be overlooked however that in the present generation there are probably more people inclined to ridicule the idea of the possibility of science in the social fields than there are to defend it. The physical sciences have been greatly developed and have won general recognition and acclaim; certain

biological sciences, particularly as represented by the field of medicine, follow upon a long history of human trust and dependence, and it is possible that in the popular mind they may be held in esteem somewhat beyond their achievement; but the social sciences yet face the task of bringing to the attention of the average person the thought that there is some possibility of discovering and in part controlling the antecedents of phenomena in social life. Promising research however is going forward, particularly under the stimulating and protecting influence of universities, where the social sciences are being nurtured until they are mature enough to command general respect. This is taking place because a few adventuring intellects have disregarded the confines of the established fields and have entered unexplored areas, making the somewhat grand assumption that events in these new fields also can be characterized in terms of law.

Another aspect of this mind-object relationship of equal importance is the assumption that the science which will be developed will comprehend the most significant aspects of the field. We are somewhat used to thinking of science as being largely concerned with laws. But we must recognize that laws as we know them are conceptual, and come into existence in large part because of the tendency of the human mind to abstract, standardize, and generalize. While many of the laws that have been promulgated have proved quite useful, they represent after all only one way of looking at the universe. There may be other ways which, when discovered and developed, will prove to be more useful and perhaps still more in agreement with what then appear to be "the facts." Even in the physical sciences, the conception of the nature of laws which obtained a few years ago is now giving place to a concept of somewhat broader functionality, and some physicists have gone so far as to express doubt about the prevalence of law in the microcosmic world. It is just possible that science in the social fields will find its broadest usefulness through the formulation of criteria somewhat different from those which have obtained for the physical sciences.

Our constantly enlarging perspective of the nature of the universe, and the light that this throws on our understanding of each particular area, make evident the partial character of our present knowledge about every particular thing. Even if we granted that our perception were perfect, and that laws as we think of them were rather discrete elements of the universe, it would still be necessary for knowledge, during any

contemporary period, to rest upon certain contingencies arising from its incompleteness. The existent world is not subdivided into independent compartments, and new developments in a field may require at any time a reorientation of the basic principles either of that field or of some related field. Many of the fundamental concepts and the broadest generalizations of the physical sciences have, during the present generation, undergone material modification; and the work of this generation will be refined further by the next. Successively as the oldest postulates are found not to be inconsistent with the accumulated insights of ages of critical inquiry, they approach more nearly the realm of absolute truth. Such truth, however, is to be regarded as a limit which is to be approached as an asymptote over an indefinite period of time. Meanwhile our conclusions must be maintained within the limits of reasonable reserve because of varying degrees of probability of subsequent modification or invalidation.

The ever expanding borders of science not only keep alive a certain possibility of old laws being superceded or subsumed by new ones, but, in bringing new areas into purview, they also add new increments of contingency as the older areas slowly approach stability. While therefore we are constantly penetrating deeper into the intricacies of nature, and undoubtedly making substantial progress in bringing our whole system of concepts into greater consistency and into closer parallelism with the external world, it is not to be expected that we shall ever completely free our knowledge of uncertainties. Each generation finds itself capable of devising ways for objectifying and quantifying various characteristics of hitherto untouched realms, and in so doing the cycle of progress is repeated. New assumptions are continually created, either incidentally or purposely, by the exploring scientist who sets about to expand the limits of his field. Thus science must ever remain, at least in some areas, avowedly contingent.

Having reached and considered the most fundamental and persistent elements of question in science, we may return to a brief statement of the ground covered. Four types of assumptions which enter into research and leave some trace of uncertainty in its conclusions have been reviewed. On the first level are the technical assumptions which embody approximations and conditional relationships of a somewhat narrow sort, many being of considerable importance in refined work, but most being of negligible effect in exploratory work.

In the second group were placed the principal decisions which have to be made in the prosecution of research, being concerned with the various aspects of methodology, and dependent largely upon the astuteness and the integrity of the worker. These are important determiners of the quality of research. The social assumptions, comprising the third type of conditional element, are concerned with the development and the dynamics of the investigator, and arise out of the necessity for a reasonably propitious environment. They constitute factors of great force in determining the amount of research that is carried on, as well as the direction that it will take. Finally, as a fourth type, there are certain great imponderables which underly all knowledge and all science. Perhaps remaining forever beyond the limits of unassailable certainty, these are disposed of for practical purposes, by being recognized as postulates, and upon these the entire superstructure of serious work is reared.

In thus calling attention to elements of possible and unavoidable uncertainty in knowledge and in science, there has been no thought of derogating from the unique attainments of both practical and pure research. Rather, the purpose has been to consider the nature of assumptions and the role they play in man's unceasing effort to push back his great intellectual frontier. It is possible that a consideration of assumptions in all their aspects will lead to a new regard for them. They are commonly thought of as disquieting elements in an otherwise untroubled, scientific serenity. They make all potential certainties somewhat contingent. They surround promising results with doubt. They frequently establish a gap between simplified theory and complex reality. But these represent only one of the roles which assumptions play. In the production of research, assumptions are dynamic. They are the means which man has developed for making progress in spite of recognized limitations. If all research is rooted in assumptions, then it is fair to say that all science has become possible by virtue of assumptions. Instead of waiting idly for the full and final understanding of the universe to burst forth in ultimate perfection upon the consciousness of mankind, scientists have assumed basic premises and have forged ahead, studying the universe little by little. They have devoted themselves to their cause, believing that practical benefits would accrue from their unremitting inquiry even though their labors might not provide ultimate answers to some of their questions—and the world has been well rewarded.

The role of assumptions in research is therefore dual: In the interpretation and utilization of research results, assumptions are restrictive; in the steady march of science from man's almost complete ignorance to his almost complete understanding, assumptions bridge unavoidable gaps and release the energies of mankind for productive work in chosen fields.

# COMPARATIVE PUPIL GAINS UNDER INDIVIDUAL CONFERENCE AND CLASSROOM INSTRUCTION

GEORGE W. HARTMANN

The Pennsylvania State College

## I. WHAT IS THE EFFECT OF CLOSE PUPIL-TEACHER CONTACT UPON PUPIL ACHIEVEMENT?

It has long been assumed that one of the main, if not the main, educational loss resulting from large classes is the diminution of personal contact between student and teacher. No one doubts that a reduction in pupil-instructor intimacy normally and inevitably accompanies an increase in the size of the group to be taught—a fact which has been deplored by most lay and professional observers. It is also commonly held that an absence of individual attention must lead to lowered achievement. Recent studies on class size have made this belief less tenable than formerly, but the conviction nevertheless persists.

Many teachers have noticed that students' morale appears to be raised by the instructor's friendly participation in their activities and that the good-will thus established seems to exercise a facilitating influence on the amount and quality of pupil performance. Unfortunately, while much of the student's yearning to be the beneficiary of the instructor's sociability involves a healthy desire to communicate experiences, it is easily perverted into "hand-shaking" and "chiselling"—practices severely frowned upon by both student-body and teaching staff because of the recognition that a pupil's mark may be influenced favorably but illicitly by such familiarity.

A neglected aspect of this problem is the fact that not all pupil-teacher contacts have positive educational value. The clerical routine of college registration and mechanical advisory functions are rarely inspiring, and may even hinder rather than foster the best conditions for mental growth. Moreover, no teacher can get along equally well with every pupil and *vice versa*, so that a certain measure of selectivity (attraction and avoidance) in these personal relations is inevitable. A final consideration is that a few minutes' intercourse with a gifted or renowned personality may be more rewarding than days spent in the presence of mediocrity.

In the study here reported, the only type of contact involved was strictly academic, *i.e.*, the interview discussions were limited to matters pertaining to the content of the course. Generalized chatter was rigorously excluded. The situation was so arranged that the student was impressed with the necessity for a sober and business-like discharge of his semester obligations. While the conversations were not exclusively in a heavy vein, they were normally serious in tone.

Specifically, will a student do better work in an elementary psychology course when he is given a brief office conference with the instructor in lieu of attending one of the class lecture or recitation periods? Does this slight reduction in the mental and social distance between the student and instructor affect the pupil's ultimate standing on objective tests and examinations given during the course?

## II. TREATMENT OF THE INTERVIEW AND THE NON-INTERVIEW GROUPS

In the fall term of 1933, the author had more than one hundred students in a class in introductory psychology, in which Murphy's *General Psychology* was used as a text. For the purposes of this experiment, the class was divided into two groups of fifty-six students each. These groups were previously matched on the basis of their average college grades. Group *A*, the control group, was required to attend the class regularly three times weekly and to participate in all the lectures, recitations and demonstrations. Group *B*, the experimental group, was forbidden to come to class for the third hour, but instead directed to make appointments for a twenty-minute conference with the instructor monthly. Three such conferences were held with each student in Group *B* during the term. During each individual session a discussion of the student's past and present reading in the course occurred and a plan for some original form of essay writing or experimental report was developed. This yielded an imperfect sort of tutorial system—a compromise adapted to the exigencies of mass instruction in a typical American state university.

Under the scheme, Group *B* was deprived of one full hour's recitation or lecture weekly, but received in return the privilege of three brief personal conferences which were denied to Group *A*. Apart from this, all work done by the two groups was of the same nature and amount. Consequently, the specific question to be solved by this arrangement was: What different levels of mastery of subject-matter result from formal group learning as distinguished from individual consultations?

### III. MEASURING THE COMPARATIVE ACCOMPLISHMENT

From the description of the teaching procedures given above, it will be seen that the students understood that the principal emphasis was on the scholarly goals to be reached and not on the tactics of attaining them. The appraisal of the relative merit of the two devices was made from the standpoint of "Does the student know the material?" At the present stage of mental measurement technique, these determinations of "knowing" had to rest upon the results obtained by administering three objective (= True-False and Multiple Choice) tests at the end of each third division of the semester. These examinations were composed exclusively of material derived from Murphy's text, which both groups were equally well prepared to take so far as opportunity to read and to understand it was concerned.

The first two tests consisted of one hundred items each; the third and last, being a final examination, consisted of one hundred ninety items. All the necessary statistical data are contained in Table I which is self-explanatory to the technically-initiated reader.

The quantitative record clearly favors the control group. On the first test, the chances are 90 in 100 in favor of it, on the second 87 in 100, and on the third 98.9 in 100. The two minor bases of comparison which are appended in the column at the right half of Table I confirm the uniform inferiority of the experimental groups. Only in the case of the term papers is there any evidence of greater achievement by the experimental group. It is, therefore, highly probable, even if not wholly certain, that our *A* group derived more tangible gains from the term's work than did the *B* group, despite its lack of the putative benefits of face-to-face relations with the instructor.

### IV. DISCUSSION AND CONCLUSION

On the surface, these findings offer no support to the clamor for more individualized teaching on the college level, at least so far as elementary or "orientation" courses are concerned. Merely knowing the teacher as such and talking with him about pertinent intellectual affairs exercises no visible influence upon the learner's willingness or ability to master the facts, technique or viewpoint of introductory psychology. To be sure, these benefits may and probably do occur when the personal contacts are *in addition* to the regular activities of the classroom, but when they are used as a *substitute* for a part of

TABLE I.—PUPIL GAINS IN ELEMENTARY PSYCHOLOGY UNDER CONDITIONS OF GROUP LEARNING AND PARTIALLY INDIVIDUALIZED INSTRUCTION

| Group | Test I, $N$ = 54 pairs, final $r_{AB}$ = .265 | | | Test II, $N$ = 54 pairs, final $r_{AB}$ = .409 | | | Test III, $N$ = 56 pairs, final $r_{AB}$ = .329 | | | Multiple choice question | | Discussion (essay) questions | | Term papers | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | $SD_{av.}$ | Mean | SD | $SD_{av.}$ | Mean | SD | $SD_{av.}$ | Mean | AD | Mean | AD | Mean | AD |
| (*B*) Experimental.... | 66.04 | 7.32 | .996 | 64.24 | 8.46 | 1.15 | 60.96 | 6.75 | .902 | 16.00 | 3.57 | 14.45 | 5.29 | 2.05 | .37 |
| (*A*) Control..... | 66.98 | 7.86 | 1.07 | 65.70 | 8.52 | 1.16 | 63.39 | 7.38 | .987 | 17.36 | 6.29 | 16.02 | 4.41 | 1.85 | .31 |
| | Diff. = .94 | $SD_{diff.}$ = .736 | Diff. / $SD_{diff.}$ = 1.28 | Diff. = 1.46 | $SD_{diff.}$ = 1.25 | Diff. / $SD_{diff.}$ = 1.17 | Diff. = 2.43 | $SD_{diff.}$ = 1.096 | Diff. / $SD_{diff}$ = 2.22 | Diff. = 1.36 | | Diff. = 1.57 | | Diff. = .20 | |

them (as in this study), they fail to produce equivalent results. This does not deny that genuine human values may occasionally result from these experiences, but they are not to be detected by any existing instruments.

One must recognize the time factor as the crucial determiner of the relative efficiency of these two policies. From the standpoint of the student's schedule, obviously sixty minutes spent in three short semester conferences are not equivalent to fifteen hours of group lectures. It could be plausibly argued that had the difference of fourteen hours been spent in such personal learning interviews, the conference group would ultimately have forged ahead of the others. Such a proposal, however, is clearly impracticable from the standpoint of the best investment of the instructor's time under normal administrative conditions. All that this study can be said to demonstrate is that there is less discernible student progress *per unit of teaching time and effort* under the individual conference plan than under group instruction.

This poor showing of the conference procedure was contrary to the writer's expectations, for he had earlier called attention to some of its theoretical advantages.[1] While these advantages (particularly on the side of pupil expressiveness) may still hold, if they can also be approximated by normal classroom operations, then a more extensive use of the interview device is doomed on the ground of its inordinate time demands. After all, the reason that group instruction evolved was the common-sense recognition that if a teacher is to communicate an idea he can do that for a dozen or one hundred as readily as for one—in fact, he may exhibit a more workman-like job with the large audience because of a recognition that more is at stake.

Possibly the avowed craving of many undergraduates for more teacher contact is simply a form of self-assertiveness—a way of bringing "my" problems and "my" ideas to the front. It probably wounds a student to be neglected, but if the instructor tries to remedy this by spreading his personality systematically and evenly over all, he still fails to meet the special wants of those individuals who can only be satisfied by distinct social prerogatives not shared by the rest of their associates. In short, a little personal treatment may make stu-

---

[1] George W. Hartmann: "The interview as a research and teaching device." *Journal of Applied Psychology*, Vol. XVII, 1933, pp. 205–211.

dents a bit happier,[1] but it demonstrably fails to raise their scholarship above the group level. Perhaps this paradoxical statement sums up best the essential conclusions of this report.

## V. SUMMARY

Fifty-six equated pairs of students in elementary psychology were the subjects of a parallel-group experiment in which one-half served as "controls" by attending recitations and lectures regularly thrice weekly, and the other half constituted an experimental set which was allowed private conferences with the instructor in place of the third weekly class period. In five out of six measures of pupil attainment (four of which were objective tests), the "interviewed" group did more poorly. It is suggested that the alleged benefits of greater student-teacher contacts are either fictitious or effective in some other way than in the overt manifestations of scholarship.

---

[1] This seems to be a legitimate inference from a little inquiry concerning the students' liking for the prescribed interviews which was made anonymously at the end of the course. Out of one hundred available responses, seventy-three declared the conferences highly helpful and recommended their continued use in the future for other students; twenty declared them "somewhat helpful"; three insisted that they were of no particular value; and four gave no answer.

# A FORMULA FOR FINDING THE AVERAGE CORRELATION OF ANY ONE VARIABLE WITH THE $(n-1)$ OTHER VARIABLES WITHOUT SOLVING ANY OF THE INDIVIDUAL CORRELATIONS

HAROLD A. EDGERTON

Ohio State University

Recently a formula has been published[1] by means of which the average of all the possible intercorrelations of $n$ variables could be obtained, without having to solve any of the individual correlation coefficients. A companion formula is here presented by means of which one may obtain the average of the correlations of any one variable with $(n-1)$ other variables, the mean of a column of a table of intercorrelations. Such a function would be of use in the solution of partial regression weights by an approximation method recently developed by Wherry.[2]

The mean of the correlations of, say variable 1 with variables 2 to $n$ is

$$\bar{r} = \frac{1}{n-1}(r_{12} + r_{13} + \cdots + r_{1n}) \tag{1}$$

Expressing the correlation coefficients as functions of standard scores,

$$r_{12} = \frac{\Sigma z_1 z_2}{N}, \text{ where } z_1 = \frac{x_1}{\sigma_1} = \frac{X_1 - M_1}{\sigma_1} \text{ and } z_2 = \frac{x_2}{\sigma_2} = \frac{X_2 - M_2}{\sigma_2}$$

Equation 1 may be rewritten,

$$\bar{r}_1 = \frac{1}{N(n-1)}\left[\sum\left(\frac{X_1 - M_1}{\sigma_1}\right)\left(\frac{X_2 - M_2}{\sigma_2} + \frac{X_3 - M_3}{\sigma_3} + \cdots + \frac{X_n - M_n}{\sigma_n}\right)\right] \tag{2}$$

---

[1] Edgerton, Harold A., and Toops, Herbert A.: "A formula for Finding the Average Intercorrelation Coefficient of Unranked Raw Scores Without Solving Any of the Individual Intercorrelations." *Journal of Educational Psychology*, February, 1928, pp. 131–138.

[2] Wherry, Robert J.: *An Approximation Method of Obtaining Partial Regression Weights.* An unpublished manuscript.

Expanding and collecting like terms,

$$\bar{r}_1 = \frac{1}{N(n-1)}\left[\sum\left(\frac{X_1}{\sigma_1}\right)\cdot\left(\frac{X_2}{\sigma_2}+\frac{X_3}{\sigma_3}+\cdots+\frac{X_n}{\sigma_n}\right)-\sum\left(\frac{X_1}{\sigma_1}\right)\cdot\left(\frac{M_2}{\sigma_2}+\frac{M_3}{\sigma_3}+\cdots+\frac{M_n}{\sigma_n}\right)-\frac{M_1}{\sigma_1}\cdot\left(\frac{\Sigma X_2}{\sigma_2}+\frac{\Sigma X_3}{\sigma_3}+\cdots+\frac{\Sigma X_n}{\sigma_n}\right)+\frac{NM_1}{\sigma_1}\cdot\left(\frac{M_2}{\sigma_2}+\frac{M_3}{\sigma_3}+\cdots+\frac{M_n}{\sigma_n}\right)\right] \quad (3)$$

The second and fourth terms in the brackets drop out, since they are of equal value and of opposite sign. Noting that $\Sigma X = NM$, equation 3 may be written,

$$\bar{r}_1 = \frac{1}{N(n-1)}\left[\sum\left(\frac{X_1}{\sigma_1}\right)\cdot\left(\frac{X_2}{\sigma_2}+\frac{X_3}{\sigma_3}+\cdots+\frac{X_n}{\sigma_n}\right)-\frac{\Sigma X_1}{N\sigma_1}\left(\frac{\Sigma X_2}{\sigma_2}+\frac{\Sigma X_3}{\sigma_3}+\cdots+\frac{\Sigma X_n}{\sigma_n}\right)\right] \quad (4)$$

Now let

$$S\frac{X}{\sigma} = \left(\frac{X_1}{\sigma_1}+\frac{X_2}{\sigma_2}+\cdots+\frac{X_n}{\sigma_3}\right)$$

and

$$S\left(\frac{\Sigma X}{\sigma}\right) = \left(\frac{\Sigma X_1}{\sigma_1}+\frac{\Sigma X_2}{\sigma_2}+\cdots+\frac{\Sigma X_n}{\sigma_n}\right)$$

Substituting in equation 4

$$\bar{r}_1 = \frac{1}{N(n-1)}\left[\sum\left(\frac{X_1}{\sigma_1}\right)\cdot\left(S\frac{X}{\sigma}-\frac{X_1}{\sigma_1}\right)-\frac{\Sigma X_1}{N\sigma_1}\left[S\left(\frac{\Sigma X}{\sigma}\right)-\frac{\Sigma X_1}{\sigma_1}\right]\right] \quad (5)$$

The general formula for the average intercorrelation of any variable $i$, with the $(n-1)$ other variables is written by substituting the subscript $i$ for the 1 in equation 5.

$$\bar{r}_i = \frac{1}{N(n-1)}\left[\sum\left(\frac{X_i}{\sigma_i}\right)\cdot\left(S\frac{X}{\sigma}-\frac{X_i}{\sigma_i}\right)-\frac{\Sigma X_i}{N\sigma_i}\left[S\left(\frac{\Sigma X}{\sigma}\right)-\frac{\Sigma X_i}{\sigma_b}\right]\right] \quad (6)$$

Formula 6 may be rewritten in better form for computation

$$\bar{r}_i = \frac{1}{N(n-1)}\left[\sum\left(\frac{X_i}{\sigma_i}\cdot S\frac{X}{\sigma}\right)-\frac{\Sigma X_i^2}{\sigma_i^2}-\frac{\Sigma X_i}{N\sigma}\left(S\frac{\Sigma X}{\sigma}-\frac{\Sigma X_i}{\sigma_i}\right)\right] \quad (7)$$

As an example of the operation of equation 7 let us take an example of only four cases and, say five variables and determine $\bar{r}_2$. It is necessary to have all the original gross scores or step scores. These are shown in Table I.

1. Find the standard deviation of each variable. The gross formula score

$$\sigma = \sqrt{\frac{N\Sigma X^2 - (\Sigma X)^2}{N^2}}$$

TABLE I.—ORIGINAL SCORES

| Persons | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|
| *A* | 3 | 1 | 9 | 7 | 2 |
| *B* | 5 | 5 | 4 | 1 | 4 |
| *C* | 7 | 4 | 2 | 4 | 5 |
| *D* | 8 | 3 | 2 | 6 | 4 |
| $\Sigma X$ | 23 | 13 | 17 | 18 | 15 |
| $\Sigma X^2$ | 147 | 51 | 105 | 102 | 61 |
| $\sigma$ | 1.92 | 1.48 | 2.86 | 2.29 | 1.09 |

TABLE II.—GROSS STANDARD SCORES

| Persons | $\frac{X_1}{\sigma_1}$ | $\frac{X_2}{\sigma_2}$ | $\frac{X_3}{\sigma_3}$ | $\frac{X_4}{\sigma_4}$ | $\frac{X_5}{\sigma_5}$ | $S\frac{X}{\sigma}$ | $\frac{X^2}{\sigma} \cdot S\frac{X}{\sigma}$ |
|---|---|---|---|---|---|---|---|
| *A* | 1.56 | .68 | 3.15 | 3.06 | 1.83 | 10.28 | 6.9904 |
| *B* | 2.60 | 3.38 | 1.40 | .44 | 3.67 | 11.49 | 38.8362 |
| *C* | 3.65 | 2.70 | .70 | 1.75 | 4.59 | 13.39 | 36.1530 |
| *D* | 4.17 | 2.03 | .70 | 2.62 | 3.67 | 13.19 | 26.7757 |
| Sums | 11.98 | 8.79 | 5.95 | 7.87 | 13.76 | 48.35 | 108.7553 |

is the most usable as it gives as a by-product certain quantities very useful for checking.

2. Divide each gross score in Table I by the $\sigma$ of the distribution to which it belongs. These "gross standard scores" should be recorded in the proper column of Table II. If there are many cases, a table of the various possible scores divided by $\sigma$ is best to use and very easy to construct.

3. Sum these $X/\sigma$ scores both vertically and horizontally. The sum of the column sums and the sum of the row sums must be identical (48.35).

4. The accuracy of the gross standard scores, $X/\sigma$, may be checked with the corresponding column sums of Table II. For variable 1, $\Sigma X = 23$, $\sigma = 1.92$. The ratio 23/1.92 should be about equal to the sum 11.98, etc. A small discrepancy is permitted, due to dropping of fractions in the $X/\sigma$ scores.

5. The row sums are $S\frac{X}{\sigma}$ values.

6. Divide the sum of the $S\frac{X}{\sigma}$ values by $N$ to obtain $S\frac{M}{\sigma}$, here 12.09. This may be checked by dividing each column sum by $N$, and adding the quotients. Each such quotient is a $M/\sigma$ value.

7. The average of the correlations $r_{12}$, $r_{23}$, $r_{24}$, $r_{25}$, is wanted. That is, the average of the correlations of variable 2 with variables 1, 3, 4, and 5. Therefore multiply each $S\frac{X}{\sigma}$ by the corresponding $\frac{X_2}{\sigma_2}$, recording in the column $\left(\frac{X_2}{\sigma_2} \cdot S\frac{X}{\sigma}\right)$.

8. Sum the column of values obtained in 7 obtaining

$$\sum\left(\frac{X_2}{\sigma_2} \cdot S\frac{X}{\sigma}\right) = 108.7753.$$

9. Substituting the values obtained in equation 7,

$$\bar{r}_2 = \frac{1}{4(5-1)}\left[108.7753 - \frac{51}{2.1904} - \frac{8.79}{4}(48.35 - 8.59)\right]$$

Solving

$\bar{r}_2 = \frac{8.5190}{16} = .532$, the average of the correlations of variable 2 with variables 1, 3, 4, and 5.

# THE INFLUENCE OF COLLEGE TRAINING UPON SUCCESS AFTER COLLEGE AS MEASURED BY JUDGES' ESTIMATES

W. R. VAN VOORHIS AND A. C. MILLER

Pennsylvania State College

## NATURE OF THE PROBLEM

The question of the influence of college training upon success of individuals after graduation has received little attention. The extent to which participation in various types of college activities contributes to later life achievement has been more conjectural than scientific. Some studies have been made, but few have been sufficiently analytical to determine the extent to which college factors vary in their contribution toward later life success. Smith pointed out that the high ranking graduate has a better chance to succeed in life than has his lower ranking classmate.[1] His study was confined to estimation of success in terms of scholastic achievement. In a study conducted at State Teachers College, Slippery Rock, Pennsylvania, Wagenhorst found practically no correlation between intelligence as measured by the Otis test, and teaching success as measured by superintendents' ratings.[2] He found a slight positive correlation ($.23 \pm .05$) between practice teaching ratings and first year of service ratings.

Other studies have been conducted but none of them seem to give satisfactory replies to a great number of questions which should be the concern of every college. Does the student who has attained high proficiency in his college work continue to be counted among those whom we look upon as successful citizens? Does the type of work taken in college have any bearing upon the student's chances for later success? What, if any, are the relationships between participation in athletics, music, school publications, or social fraternities, and position in society when these direct influences have been removed? Does the student who excels in mathematics or languages acquire certain attributes which tend to place him higher in the estimation of his fellow men than the one who has concentrated on the social sciences?

[1] Smith, Hugh A.: "College Records and Success in Life." *Education*, Vol. XLVII, No. 9, pp. 513.

[2] Wagenhorst, L. H.: "The Relationship Between Rating of Student Teachers in College and Success in First Year of Teaching." *Educational Administration and Supervision*, Vol. XVI, No. 4, April, 1930, pp. 249.

These and many other similar questions have been the chief stimulation of initiation of this study.

The authors have attempted a step in the direction of answering some of these questions. No claim is made that definite answers will be given. Rather, it is to be hoped that more extensive investigation be carried on in the future in order to ascertain more reliably the extent to which differences in college training influence the later life success of individuals, and to establish a technique and procedure which may prove definitely applicable to problems of this nature.

## STATEMENT OF THE PROBLEM

The present study attempts to determine the influence of certain college factors upon the later success of individuals as measured by the estimates of persons who have known many of these individuals since their graduation from college. Data concerning the class which graduated from the Pennsylvania State College in the year 1928 serve as the basis of the study. Most of the information concerning the college activities of the members of that class was obtained from the offices of the Registrar of that institution. College grade point averages, Carnegie Foundation Test Scores, DeCamp Psychological Test Scores, grades in mathematics, languages and social studies, as well as information concerning participation in athletics, publications, music, advanced military drill, and social and honorary fraternities were obtained from that source. The list of Honor Students was taken from the Commencement program of June, 1928.

## METHOD OF ANALYSIS

In the statistical treatment of data, the tetrachoric correlation technique was employed in most cases. Three different tetrachoric formulae were used depending upon the way in which the two correlated distributions were divided. When both correlated distributions were divided at the means Pearson's cosine-$\pi$ formula was used. If either or both of the distributions were divided at some point other than the mean, Pearson's general tetrachoric formula was used. In the cases of mutilated distributions of widely separated dichotomies, a formula based on Pearson's general formula and developed by Dr. C. C. Peters and W. R. Van Voorhis was employed.[1] This formula makes

[1] This formula is developed in a book on advanced statistics to be published shortly by these authors.

possible the computations of $r$'s when data are available for only the extreme tails of distributions, as well as instances in which these tails are unequal. The bi-serial method served wherever one distribution was complete and the other mutilated about the mean. In all cases in which paired scores were present for both distributions, the Pearson Product-Moment formula was used.

Partial correlation coefficients were obtained in some cases to show the extent of agreement between certain factors with certain other factors held constant. No particular formula was singled out for the computation of the intercorrelations for this purpose, since the $r$'s obtained by all of the above mentioned formulae have the same meaning. A limited number of the intercorrelations tested was selected in the computation of the partial $r's$ because of low zero-order $r's$ in some cases, and because of small numbers of cases involved in others.

## THE CRITERION OF SUCCESS

Judges' estimates constituted the criterion of success. This measure was used for at least two reasons: In the first place, in the absence of more objective measures, this criterion seemed the best available; secondly, it is probably as reliable a measure as could be obtained from any source. That judges' estimates, when taken over a long period of time, or when many in number, are a quite reliable index of achievement, is coming into recognition. Smith made use of judges' estimates and he states that his answers made it clear that an intelligent public will generally agree on what constitutes success.[1] Studies conducted at the Pennsylvania State College have yielded reliability coefficients of judges' estimates as high as .99.[2] Indeed, it is true that estimates, as a measuring instrument, are in many cases superior to many so-called objective tests. Kelly states: "Best prediction of success in high school is obtained from elementary school marks, next best prediction from teachers' ratings, and the least valuable from tests."[2] It is not, therefore, without experimental justification that estimates have been used to measure post-college success in this study.

On the basis of estimates made by sixty-seven individuals who have known certain members of the graduating class of 1928 since

[1] *Loc. cit.*, p. 1.

[2] Swab, James: *The Reliability and Validity of Pupil Estimates as a Measuring Tool*, Penn State Studies in Education, Part III, 1933.

that year, a portion of the total number of five hundred eighty-two graduates was divided into two groups, which we have called the "successful" group and the "unsuccessful" group. These groups were obtained as follows: The complete list of the names of the five hundred eighty-two graduates was presented to each of the sixty-seven judges with the instructions to place in one group those whom he considered *very successful* in their lives since graduation from college, and in the other group those whom he considered *very unsuccessful* during the same period. This does not mean that each of the sixty-seven judges rated each individual. The number of judgments per individual ranged from one to twenty-three for the "successful" group and from one to seventeen for the "unsuccessful" group. The mode for each group was seven. It was assumed that variations in number of judgments per individual were indicative of an individual's ability to be placed at different range levels in either of the two groups. Therefore, in order to insure classification in the extreme tails of the success area, the demand was made that an individual receive seven or more nominations in order to be placed in one of the two groups. No attempt was made to define the term "success." The placement of an individual in one category or the other was left entirely to the judges. In this manner, under the assumption of normality of distribution of those elements or abilities which constitute "success," it was possible to obtain nominations of individuals in both the upper and the lower portions of the total "success" area. This procedure has practical advantages, not only from the standpoint of statistical treatment, but also because of its ease of classification. Kriner states: "For practical purposes it appears reasonable that these two groups can be contrasted better, and fewer will be placed in the wrong category than could a group where the large central portion of the distribution would have to be placed either in one or the other category because of a very meager difference."[1]

The number of nominations to the "successful" group ranged from one to twenty-three, while that of the "unsuccessful" group ranged from one to seventeen. The mode for each group was seven. These variations in number of judgments per individual were interpreted to be indicative of variations in abilities among individuals to be placed at different range levels in either portion of the total "suc-

---

[1] Kriner, H. L.: *Pre-Training Factors Predictive of Teacher Success.* Penn State Studies in Education, No. 1, 1931.

cess" area tested. Of the total of five hundred eighty-two graduates, one hundred fifty-one or 25.9 per cent were named "successful" and one hundred six, or 18.2 per cent were named "unsuccessful." Thus, a representative sampling of the most successful and the least successful of the total graduating class of 1928 was obtained instead of the relative position according to success for each member of the group.

## FINDINGS OF THE STUDY

The relationships of various college factors to success are given in the form of coefficients of correlation. These are presented in Table I.

TABLE I.—COEFFICIENTS OF CORRELATION BETWEEN THE SUCCESS OF COLLEGE GRADUATES AND CERTAIN OTHER FACTORS

| SUCCESS CORRELATED WITH | r |
|---|---|
| 1. DeCamp psychological test scores | .187 |
| 2. College grade point averages | .331 |
| 3. Carnegie foundation test scores | .278 |
| 4. Mathematics grade point averages | .182 |
| 5. Language grade point averages | .202 |
| 6. Social studies grade point averages | .319 |
| 7. Participation in music | .080 |
| 8. Participation in publications | .113 |
| 9. Participation in advanced military drill | .580 |
| 10. Membership in honorary fraternities | .326 |
| 11. Membership in social fraternities | .101 |
| 12. Participation in athletics | .084 |
| 13. Scholastic honors | .337 |

The highest correlation with success is that of advanced military drill (.580). Other noticeable correlations are with college grade point averages (.331), membership in honorary fraternities (.326), and scholastic honors (.337). It is interesting to note that of all the academic items tested, social studies and success correlate highest (.319). There is seen to be little correlation of success with music (.080) and no more with athletics (.084). The correlation of Carnegie foundation test scores with success (.278) is not as high as those factors which represent scholastic achievement.

The ninety-one intercorrelations of all factors tested in relation to success are presented in Table II.

The Decamp test scores correlate positively with all other factors; highest with social studies (.497), and lowest with music (.113). College grade point averages correlate negatively with but the single

TABLE II.—INTERCORRELATIONS OF ALL COLLEGE FACTORS AND SUCCESS

| | DeCamp | College | Carnegie | Mathematics | Language | Social studies | Music | Publications | Military drill | Honorary fraternities | Social fraternities | Athletics | Honors |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| College | .356 | | | | | | | | | | | | |
| Carnegie | .399 | .269 | | | | | | | | | | | |
| Mathematics | .411 | .505 | .368 | | | | | | | | | | |
| Language | .418 | .641 | .361 | .417 | | | | | | | | | |
| Social studies | .497 | .510 | .554 | .218 | .353 | | | | | | | | |
| Music | .113 | .142 | .028 | −.142 | .000 | −.225 | | | | | | | |
| Publications | .335 | .001 | .271 | −.083 | .059 | .130 | .349 | | | | | | |
| Military drill | .257 | .148 | .185 | −.150 | .112 | −.075 | .119 | .424 | | | | | |
| Honorary fraternities | .400 | .366 | .259 | .323 | .372 | .338 | .301 | .637 | .588 | | | | |
| Social fraternities | .460 | .049 | −.177 | −.169 | .126 | −.311 | .227 | .342 | .104 | .208 | | | |
| Athletics | .168 | −.077 | .271 | .000 | .019 | −.173 | .215 | .015 | .145 | .123 | .388 | | |
| Honors | .288 | 1.000 | .387 | .657 | .675 | .415 | .011 | −.035 | −.066 | .417 | −.227 | −.096 | |
| Success | .187 | .331 | .278 | .182 | .202 | .319 | .080 | .113 | .580 | .326 | .101 | .084 | .337 |

item of athletics (−.077). Its highest positive correlation is with languages (.641). There is seen to be practically no agreement between college grade point averages and publications or membership in social fraternities. Carnegie foundation test scores correlate positively with all other items except with social fraternity membership (−.177). Mathematics correlates positively with college grade point average (.505), Carnegie test scores (.368), languages (.417), honorary fraternities (.323), and honors (.657). Slight negative correlations are found between mathematics and music, publications, advanced military drill and social fraternities; no correlation with athletics. Languages and social studies correlate with other college factors in about the same manner and to about the same degree as does mathematics. Music and publications correlate but little with other factors. Honorary fraternity membership correlates to a noticeable degree with all others except with athletics. The correlations of social fraternities with other factors vary from (−.311) with social studies to (.460) with Decamp test scores. There is no significant correlation between athletics and other factors except possibly with Carnegie scores (.271) and with social fraternities (.388). Honors correlate with all other factors in about the same manner as do college grade point averages. Advanced military drill, honors and grade point averages correlate highest with success, while

music, publications, social fraternities and athletics correlate lowest with it.

Certain of the items of Table II were excluded from a partial correlation treatment. Mathematics, languages, social studies, attainment of honors, and membership in honorary fraternities were omitted because of their overlapping with college grade point averages. Music and publications were dropped because of their generally insignificant *r*'s. Advanced military drill was omitted because of the small number of cases.

Table III displays the partial coefficients of correlation between success and DeCamp test, college grade point average, Carnegie test, social fraternity, and athletics.

TABLE III.—PARTIAL COEFFICIENTS OF CORRELATION BETWEEN SUCCESS AND A GIVEN FACTOR WITH ALL OTHER FACTORS IN THE TABLE PARTIALED OUT

| FACTORS CORRELATED WITH SUCCESS | CORRELATION |
|---|---|
| College grade point average | .516 |
| Carnegie foundation test | .223 |
| DeCamp psychological test | .127 |
| Membership in social fraternity | .081 |
| Participation in athletics | .056 |

It can be seen from Table III that college grade point average is by far the most significant of all the items tested in relation to success when the other items are held constant. The coefficient .516, for example, is the partial coefficient of correlation between the criterion success and college grade point average when Carnegie foundation test scores, DeCamp psychological test scores, membership in social fraternities, and participation in athletics were held constant. A comparison of Tables II and III shows that the partial technique has increased the *r* between college grade point average and success from (.331) to (.516). The *r* between DeCamp test and success decreased from (.187) to (.127). The Carnegie test *r* dropped from (.278) to (.223). The *r*'s for social fraternities and athletics with success were reduced from (.101) and (.084) to (.081) and (.056) respectively.

# THE MEASURABLE EFFECTS OF ZOOLOGY ON PSYCHOLOGY GRADES

A. R. GILLILAND

Northwestern University

There has been much discussion of the effects of the study of certain school subjects upon success in other subjects. In fact much of the problem of transfer of training centers around this point. However, relatively little experimental study has been made of the problem to be presented here.

This investigation attempts to measure the effect of the study of zoology (biology) on grades in general psychology. The problem is directly related to whether zoology should be made a prerequisite to courses in psychology. Of course it should be admitted at the outset that there may be certain "intangible" results accruing from the study of zoology which cannot be measured, at least, by the methods employed in this study. It may also be admitted that the study of zoology may have a value to later, more advanced work in psychology which does not become apparent in the beginning course in psychology. However, there still remains the major problem of what effects, if any, courses in zoology have on grades in general psychology.

In order to study this problem, the students in several classes in general psychology at Northwestern University were separated into those who had studied zoology, and those who had not. Only college zoology was considered, and no account was taken of high school courses in zoology. Those who had studied zoology were further classified into those who had taken one semester (up to six hours), or more than two semesters of zoology before beginning the study of psychology.

The following data was available for each student in each group. (1) The entrance grade in the American Council of Education Intelligence test: These were expressed in percentile scores making it possible to compare students who entered the University in different years. (2) The test scores on the examination in psychology which deals directly with the human organism, with special emphasis upon the nervous system and its functioning. Such grades were available because uniform weekly quizzes are given to all sections of general psychology used in this study. Although grades for two different

examinations were used, these were kept separate and provide the data for Part I and Part II of the tabular data of this study. (3) Final examination grades in general psychology. These were expressed in point scores, A = 7; A− = 6; B = 5; B− = 4; C = 3; C− = 2; D = 1; and E and F = 0.

In order to make the groups that had zoology comparable in ability with those who had studied zoology, certain cases with high percentile scores in mental alertness from the zoology group in Part II of the study, were discarded. By discarding four of these cases the two groups had almost identical mean percentile scores, as well as very similar distributions of percentile scores. In Part I the mental alertness scores and the distributions were very similar for the group with and without zoology. No cases were discarded in Part I.

The following table gives a summary of the results of the study.

TABLE SHOWING SCORES FOR STUDENTS IN PSYCHOLOGY WHO HAD AND WHO HAD NOT STUDIED ZOOLOGY

Part I—Results for 1932

| | No. of cases | Percentile MA score | | Examination scores in zoology part of general psychology | | Final grade in psychology | |
|---|---|---|---|---|---|---|---|
| | | Mean | $\sigma$ | Mean | $\sigma$ | Mean | $\sigma$ |
| With zoology..... | 71 | 50.2 | 8.5 | 56.2 | 7.7 | 4.04 | 1.5 |
| Without zoology..... | 191 | 48.6 | 30.3 | 51.8 | 8.6 | 3.79 | 1.7 |
| **Part II—Results for 1933** | | | | | | | |
| With zoology..... | 51 | 56.8 | 25.8 | 38.8 | 4.8 | 4.48 | 1.5 |
| Without zoology..... | 239 | 56.6 | 26.0 | 36.2 | 6.0 | 3.92 | 1.6 |

The significance of the difference between the scores in the weekly quiz on the human organism, the section in general psychology that is primarily biological in character was obtained. Dividing the mean difference between the scores for zoology and non-zoology students by the sigma of the difference gives 3.70 and 3.45 for Part I and II respectively. This indicates that in more than ninety-nine chances out of one hundred both these differences are significant. There is little question, therefore, but what students who have studied zoology

do better in that part of psychology which is largely zoological in character.

The students of Part I who had studied zoology for two semesters or more were separated from those who had studied zoology for only one semester. The mental alertness scores of the two groups were almost identical. The mean scores for forty-five students with two or more semesters of zoology was 57.2 $\sigma$ 5.6, and for the twenty-seven with only one semester of zoology was 53.0 $\sigma$ 6.4. By the application of the formula stated above this difference is shown to be statistically almost certainly significant; $\sigma$diff = 4.2.

Even though not so much importance can be attached to these results because they are based upon such a small number of cases, the results do compare favorably with the averages for the other groups. Seemingly even one semester of zoology has its effects upon grades in that part of psychology based upon zoological material, but the second semester of zoology has an increased value. This seems reasonable, because more of the material on human anatomy and neurology comes in the second semester of zoology.

When final grades in psychology are considered, the students who have had zoology make the better grades. In Part I there are eighty-six chances in one hundred that this difference is significant. In Part II the chances are 98.6 out of one hundred that the difference is significant. These data, so far as we can accept them at face value, indicate that the difference is significant, but not as significant as that part of the course in psychology which is primarily biological in character.

When those who had taken two semesters or more of zoology before studying psychology were compared with those who had taken only one semester, no difference was found. This is to be explained in part by the fact that many of those who had taken only one semester of zoology completed their year of zoology while they were taking their course in psychology, and received the benefits of the second semester of zoology before taking the final examination in psychology.

These results are not startling, and they are just about what common sense might lead us to expect. However, it is so often that common sense leads us astray that we need to continually check upon it by scientific investigation.

It is not the function of this article to say whether zoology should be made a prerequisite for a course in general psychology. This study has shown that zoology is helpful to the student in general

psychology. The question remains as to how helpful it should be before zoology is made a prerequisite for general psychology. This question must be answered in terms of the local institution.

If sufficient data were available it would be interesting to study the effects of zoology on more advanced courses in psychology. A further question is the effect of more advanced courses in zoology on psychology. These problems must await the accumulation of much additional data, for only with a large number of cases would the evidence be conclusive on these questions.

# A NEW METHOD OF INCREASING THE RELIABILITY OF THE TRUE-FALSE EXAMINATION

CARL IVER HOVLAND

Institute of Human Relations, Yale University

AND

JOHN C. EBERHART

Northwestern University

The conventional true-false examination is frequently criticized by students on the grounds that the statements used are often ambiguous, even when the examiner has exercised the greatest care in their preparation. It occurred to the writers that this might be one reason for the low reliability of true-false tests as compared with certain other types of objective examinations. In this paper we suggest a method which makes it possible to retain the many desirable features of the true-false test, including its simplicity of construction and grading, and which, at the same time, by reducing the possibility of ambiguity, increases the test's reliability to a point comparable to that of the multiple-choice and matching types of examination.

## THE METHOD

The fundamental change which is introduced by our method involves the underscoring of the significant or critical part of the statement, thus enabling the student to make his judgment principally on the basis of the truth or falsity of the underlined portion. A sample item will make this procedure clearer:

### Conventional Form

T F Karl Lashley found in 1926 that in rats the occipital lobes of the cerebral cortex are necessary for learning a brightness discrimination habit.

### Underscored Form

T F Karl Lashley found in 1926 that in rats the occipital lobes of the cerebral cortex are necessary for learning a brightness discrimination habit.

It can be seen from our example that when we ask, as we do in the conventional true-false test, whether the statement as a whole is true or false, we are in reality asking indirectly for answers to a large number of sub-questions, such as the following: (1) Was Lashley's first name Karl? (2) Was the investigator Lashley? (3) Was the

work done about 1926? (4) Were the experimental animals rats? (5) Is the portion of the brain required for the habit in the occipital lobe? (6) Are the occipital lobes part of the cerebrum? (7) Is it the cortical area which is essential? (8) Are the lobes necessary for learning the habit, or for some other activity, *e.g.* retaining the habit? (9) Is brightness discrimination the habit affected? When our method is employed the decision of the student is based upon the truth or falsity of but a single item. We are thus asking, in effect, only a single question, in contrast to the traditional method where, as we have seen, we are propounding a barrage of questions. In our sample item, the responsibility of the student is limited, when the significant portion is underlined, to the knowledge required to answer the eighth of the above nine questions. While all but one item of the sample statement is true, the statement as a whole is false, since the experiment to which reference is made established the fact that the occipital lobes are not necessary for learning a brightness discrimination habit, although such a habit once learned is affected by the removal of the occipital lobes. Concerning the truth or falsity of this portion of the statement there is no dispute, the required reading references having covered this factual material thoroughly. When, however, we present the same statement as a true-false item, without making clear to the student, by underlining or by some other method, the crucial portion, there is invariably a protest that the statement is ambiguous, or the point stressed too obscure.

It was believed by us that much of the ambiguity of the true-false examination is the result of requiring the student to answer, by simply putting a "true" or a "false" in front of a statement, a large battery of questions. It is extremely difficult, under such a system, to determine whether or not the student possesses the information relevant to the real point at issue. Students may correctly mark the statement "false," not on the basis of their knowledge that the crucial part of the statement is untrue, but on the basis of an incorrect belief that one of the true portions of the sentence is false. On the other hand, slight misunderstanding as to the emphasis of a question may easily cause even the most intelligent and best prepared student to mark as "false" a question which the instructor intended to be "true," in situations where the significant part of the statement is true, but where minor ambiguity suggests that part of the statement is false. We submitted to experimental test our logically derived

expectation that by emphasizing the significant part of the true-false statement the reliability of the true-false examination could be increased.

## EXPERIMENTAL PROCEDURE

Three undergraduate classes in elementary general psychology at Northwestern University were available for experimentation. In our first study we used sections one and three, with populations of ninety-two and one hundred seventy-eight, respectively. An examination was prepared covering the "Intelligence" chapters in two beginning textbooks—Gilliland, Morgan and Stevens* and Woodworth.† Sixty true-false statements were constructed, thirty-four being obtained from material in the former text and twenty-six from material in the latter. Two forms of examination were constructed. The statements were identical, but on the second form the significant or crucial part of the sentence was underlined according to the method described above. To the students taking the test in this form the following instructions were given:

*Directions.*—In each of the following true-false statements a certain word or phrase has been underlined. The underlining is to aid your analysis of the statements. Judge the truth or falsity of each statement principally on the basis of the truth or falsity of the underlined part.

Because of a possible differential difficulty between the two forms, and because the test constituted the regular weekly examination in general psychology, it seemed unwise to split classes, so one form was given to section one, and the other form to section three.

In the second study the test covered the material on "Learning" in the two textbooks used. Sections one ($N = 92$) and two ($N = 95$) were given the test, which consisted of sixty-four items. The students in the even-numbered seats were in this case given the conventional type true-false examination, while those in the odd-numbered seats were given the same statements, but with the crucial points underlined.

## RESULTS

In Table I we have compared the reliability of the standard type of true-false test with that obtained in a test containing the same

---

* Gilliland, A. R., Morgan, J. J. B., and Stevens, S. N.: *General Psychology for Professional Students.* Boston, Heath, 1930.

† Woodworth, R. S.: *Psychology.* New York, Holt, 1921.

statements, but constructed according to the method outlined by us. The number of correct answers on the even questions was correlated with the number of correct responses on the odd questions, and the reliability of the whole test computed by the Spearman-Brown prophecy formula.

TABLE I.—RELIABILITIES OF ORDINARY AND NEW-TYPE UNDERLINED TRUE-FALSE TESTS

| | Not underlined | | | Underlined | | |
|---|---|---|---|---|---|---|
| | *N* | *r* | $PE_r$ | *N* | *r* | $PE_r$ |
| Group 1 | 92 | 0.694 | 0.036 | 178 | 0.776 | 0.020 |
| Group 2 | 93 | 0.641 | 0.041 | 94 | 0.821 | 0.023 |

The critical ratios, expressing the chances of a true difference in the reliability of the two tests, are presented in Table II.

TABLE II.—DIFFERENCES IN THE RELIABILITY OF THE ORDINARY TRUE-FALSE EXAMINATION AND THE NEW-TYPE UNDERLINED FORM

| | Difference in reliability | $PE_{diff.}$ | $\frac{\text{Difference}}{PE_{diff.}}$ | Chances in one hundred that a true difference exists |
|---|---|---|---|---|
| Group 1 | 0.082 | 0.041 | 2.00 | 91.0 |
| Group 2 | 0.180 | 0.047 | 3.83 | 99.5 |

The number of correct answers given by the groups taking the conventional form of test is compared with the number of correct answers given by those taking the underlined form in Table III.

TABLE III.—MEAN NUMBER OF CORRECT RESPONSES ON REGULAR EXAMINATION AS COMPARED TO NUMBER MADE ON ALTERNATE UNDERSCORED FORM

| | Not underlined | | Underlined | |
|---|---|---|---|---|
| | Mean | $PE_M$ | Mean | $PE_M$ |
| Group 1 | 49.8 | 0.30 | 49.4 | 0.28 |
| Group 2 | 54.2 | 0.32 | 53.9 | 0.33 |

The critical ratios of these differences in mean scores are shown in Table IV.

TABLE IV.—CRITICAL RATIOS OF DIFFERENCES IN SCORE BETWEEN GROUPS TAKING CONVENTIONAL FORM AND THOSE TAKING UNDERSCORED FORM

| | Difference | $PE_{diff.}$ | $\frac{\text{Difference}}{PE_{diff.}}$ | Chances in one hundred that a true difference exists |
|---|---|---|---|---|
| Group 1..... | 0.40 | 0.41 | 0.975 | 74 |
| Group 2..... | 0.30 | 0.46 | 0.652 | 67 |

The question of whether or not the new examination has a satisfactory reliability can be determined by finding out its reliability when a fairly lengthy examination is constructed from its items. Table V gives the predicted reliability of a test of one hundred fifty items, based upon the Spearman-Brown prophecy formula:

TABLE V.—RELIABILITY OF UNDERLINED TEST FOR VARIOUS LENGTHS OF EXAMINATION (PREDICTIONS ON BASIS OF SPEARMAN-BROWN FORMULA)

| | $r$ (split half) | $r$ (entire test) | $r$ (test of one hundred fifty items) |
|---|---|---|---|
| Group 1..... | .634 (thirty items) | .776 (sixty items) | .896 |
| Group 2..... | .696 (thirty-two items) | .821 (sixty-four items) | .915 |

## DISCUSSION

Our experimental results indicate a fairly definite tendency for the tests constructed according to the method described in this paper to exceed in reliability tests made up of the same items but presented in the conventional manner. While the critical ratios do not establish complete reliability the consistency of the data strongly suggests that a true difference exists. The results of Table V indicate that with a test of greater length a reliability will be obtained which exceeds that of the majority of true-false tests (cf. *e.g.* Gilliland and Misbach[1] and Kinney and Eurich[3,4]) and equals that of other reliable objective tests (Lee and Symonds[5]).

Our test resembles somewhat the form proposed by Holmes and Heidbreder.[2] Their method involved the presentation of false state-

ments with directions to the students to underline the false portion. Satisfactory reliability is presented by their test. But our method has the advantages of utilizing both true and false questions and greater simplicity of scoring.

One might have expected that students taking the underlined form would have an advantage and would consequently make higher scores. That this was not the case is shown definitely by Tables III and IV. In fact, there is a tendency, though not statistically reliable, for the students given the analyzed form to make lower scores. A possible explanation may be that limiting the decision to a single critical point may reduce the number of cues afforded by the sentence as a whole.

The preparation of the examination impressed the authors with the desirability and importance of having a specific and critical point to each question, even though an underlining procedure is not adopted. Mere copying of statements from textbooks as material for tests is definitely discouraged by our procedure. While satisfying students may not be the fundamental objective of a testing program, we were pleased to find student opinion heartily in favor of the procedure we have outlined, despite the fact, already noted, that students taking the underlined form made no higher grades.

## SUMMARY

Two groups of students in elementary psychology, with a combined population of two hundred seventy-two individuals, were given examinations made up of true-false statements in which the significant or crucial part of the sentence was underlined. The students were instructed to make their decisions concerning the statements principally upon the basis of the truth or falsity of the underscored portion. Control groups of one hundred eighty-five students were given the same examinations without the statements being underlined. The reliability of the underlined examination was consistently higher than that of the conventional true-false test containing the same statements $\left(\frac{\text{Difference}}{\text{PE}_{\text{diff.}}} = 2.00 \text{ and } 3.83 \text{ for the two groups}\right)$. The groups taking the examination in the underlined form obtained substantially the same average scores as the control groups $\left(\frac{\text{Difference}}{\text{PE}_{\text{diff.}}} = 0.975 \text{ and } 0.652\right)$. The reliability of a test of one hundred fifty items constructed according to our method would have a reliability, according to the Spearman-Brown prophecy formula, between .90 and. 92. This value

compares favorably with the reliabilities of other types of objective tests of similar length.

REFERENCES

1. Gilliland, A. R. and Misbach, L. E.: "Relative Values of Objective and Essay Type Examinations in General Psychology." *J. Educ. Psychol.*, Vol. XXIV, 1933, pp. 349–361.
2. Holmes, G. and Heidbreder, E.: "A Statistical Study of a New Type of Objective Examination Question." *J. Educ. Res.*, Vol. XXIV, 1931, pp. 286–292.
3. Kinney, L. B. and Eurich, A. C.: "Studies of the True-False Examination." *Psychol. Bull.*, Vol. XXX, 1933, pp. 505–517.
4. Kinney, L. B. and Eurich, A. C.: "A Summary of Investigations Comparing Different Types of Tests." *Sch. and Soc.*, Vol. XXXVI, 1932, pp. 540–544.
5. Lee, J. M. and Symonds, P. M.: "New-Type or Objective Tests: a Summary of Recent Investigations." *J. Educ. Psychol.*, Vol. XXIV, 1933, pp.21–38.

# NOTE ON SCORING TESTS OF MULTIPLE WEIGHTED ITEMS

VIRGINIA I. ZERILLI[1]

Institute of School Experimentation, Teachers College, Columbia University

One of the major problems of testing programs is the expense incurred for scoring. Frequently, this expense item can be reduced considerably by making a slight change in the methods of scoring tests. Many personality tests, such as the Allport A-S Reaction Study, Grace Manson Test of Vocational Aptitudes, and the Bernreuter Personality Inventory, use a system of evaluating each item by means of negative and positive arithmetical values. For the purpose of establishing the reliability of a test before publication, this may be a useful procedure, but, when the test is actually published, a more simplified scoring method such as the one suggested in this paper is both justifiable and practical.

The writer scored about three hundred fifty Bernreuter Personality Inventories. Empirical observations made during the scoring on Bernreuter's method as compared with the one described in this paper were carefully checked.

The method suggested is to convert all values on the original scale into positive numbers. This would change the range of values for any item from (−7 to +7) to (0 to 14). The change would eliminate the need of keeping two records, one for negative and one for positive values, while scoring. There would be no necessity for using an adding machine, which Bernreuter in his manual considers indispensable for the fastest scoring. With the newer method mental addition would be more direct and faster. This method can be applied to any test using a scoring technique similar to Bernreuter's original scale.

The new scoring key is constructed as follows: For any item, add the value of the largest negative value to each of the values given, for example, an item on the Bernreuter scale may be evaluated 3 (yes), −2 (no), −6 (?). In this item, −6 is the largest negative value given. Adding 6 to 3, −2, and −6, we obtain 9 (yes), 4 (no), 0 (?). The relative value for any response to a given question remains the same. In this item, the (yes) receives five points more than (no), nine points more than (?), and so on.

[1] The writer is grateful to Professor Rudolf Pintner, Professor Otis W. Caldwell, and Dr. J. W. Wrightstone for their guidance and helpful suggestions.

In order, however, to obtain the same final rating as in the Bernreuter scale, so that we may still use the given percentile ratings and norms, we find that there is a constant negative number for each of the four different scales. This constant is the sum of each of the largest negative values for each item on the Bernreuter scoring keys. The following brief example shows how the suggested method works.

Let us assume the following to be a miniature test, and the encircled values the response of a tested person.

| Scoring according to Bernreuter scale | | | | Scoring according to suggested method | | |
|---|---|---|---|---|---|---|
| Yes | No | ? | | Yes | No | ? |
| −(3) | 3 | −1 | | (0) | 6 | 2 |
| 0 | (1) | −2 | | 2 | (3) | 0 |
| −1 | 1 | −(1) | | 0 | 2 | (0) |
| −4 | (4) | 0 | corresponds to | 0 | (8) | 4 |
| −(1) | 2 | −2 | | (1) | 4 | 0 |
| 3 | −(3) | 0 | | 6 | (0) | 3 |
| 2 | −(1) | −2 | | 4 | (1) | 0 |

| | | | |
|---|---|---|---|
| Sum of negatives | −9 | Score | 13 |
| Sum of positives | 5 | Adding the constant | −17 |
| Score | −4 | Bernreuter score | −4 |

The sum of the largest negative in each item is −17. Adding −17 to the total obtained by the suggested method we have −4, the rating obtained by the use of plus and minus values.

| | |
|---|---|
| The constant for B1N is | −294 |
| The constant for B2S is | −243 |
| The constant for B3I | −202 |
| The constant for B4D is | −245 |

The following simple experiment was made to show how much the suggested method improves upon the scoring method described by Bernreuter. Thirty Bernreuter Inventories were scored by two persons accustomed to scoring. The first person scored the set first by using the method given by Bernreuter, and then checked his work by using the new method. The second person reversed the procedure. A careful record was kept of time for scoring as well as of the number of errors made.

As a result of this experiment, the following conclusions became evident:

1. Allowing for individual variations between scores in adjustment ability, the time for scoring an Inventory by Bernreuter's method was from thirty-five to forty-five per cent longer.

2. Errors made in scoring the Inventory by using the new method for all four categories occur in approximately one out of six papers. Errors occurred in five out of six papers by using Bernreuter's method. Calculation of plus or minus signs was a frequent source of error.

3. The element of greater fatigue and strain in using Bernreuter's method was noticed, but no measurement was made.

These conclusions can be supported by a simple comparison of methods. On the original rating scale of the Bernreuter Personality Inventory, the range of values for a given item is from $-7$ to $+7$. The use of positive and negative values necessitates the keeping of two records while scoring. When adding values, the need for constant classification of plus and minus scores complicates the method, adding confusion to difficulty, as well as increasing the scoring time.

A scoring key, which, because of the very nature of a test, is long and difficult to work with, should apply as many of the simpler methods of scoring as possible. For a person who has only ten or twenty inventories to score, such details do not matter very much. It is only when he is confronted with the task of scoring fifty, one hundred, or more, that he realizes the values of a slight simplification which reduces the total time for scoring considerably. In addition, the cost of counting machines will be eliminated.

The Bernreuter Personality Inventory contains three alternative responses for each of one hundred twenty-five items. The Manson Test of Vocational Aptitudes contains five alternative responses for each of one hundred sixty items. For tests of this type, this newer method is simpler and can be used profitably. This method also suggests a simple device which can be used to convert preliminary rating scales using plus and minus values to a scale consisting of only positive values. In this way, a simple revision of the scoring key of a test before publication will prevent unnecessary complications in scoring.

# BOOK REVIEWS

ERNEST BURTON SKAGGS. *A Textbook of Experimental and Theoretical Psychology.* Boston: The Christopher Publishing House, 1935, pp. 426.

The title of Dr. Skagg's book is a little misleading. It is not a textbook for a course in Experimental Psychology—at least, not as such courses are usually constituted—but is rather designed for the introductory course. The author has succeeded well in his announced purpose of "presenting recent experimental findings and theories on a dignified scientific basis." Little of the existing body of fact and theory in psychology has been omitted, and the treatment is generally quite free from bias. It is, of course, not the ideal text, for that is something which probably no one man can write, but it is superior to many of those in existence.

The inadequacy of the book lies in its organization of material. Despite the author's avowal in the preface that "Mind or consciousness is assumed to be but an epi-phenomenon, a curious and most interesting by-product of physiological processes of the body," the first three-fourths of the book are devoted to what he calls the "facts of consciousness." Indeed, at the end of Chapter XI he says, "Our psychology has been an introspective or mentalistic psychology. Now, however, our interest shifts and we become, to a degree, objective psychologists or behaviorists"—for the last one hundred forty pages of the book! The present reviewer cannot help questioning the pedagogical desirability of such a point of view. It is difficult to see, too, the author's basis for this division, since the chapters devoted to "mentalistic psychology" include treatment of the nervous system, receptors, effectors, sensation, perception, sensory and ideational memory, thinking, emotion; while those purported to be "objective psychology" treat of unlearned motor activity, sensory-motor learning and memory the control of action, work and fatigue, sleep and dreams, individual differences, and personality. Such a division has led to the very questionable procedure of presenting part of the material on learning and memory at one place and the remainder some one hundred pages later. Even the description of the characteristics of learning curves appears at two widely separated points in the book. One might also question the logic of discussing conditioned emotions some thirty pages before the conditioned reflex is described.

The relative emphasis which a textbook should place upon topics is, perhaps, a matter of personal opinion, one author's emphasis being as warranted as another's. To this reviewer, however, the devoting of only three and a half pages to the subject of intelligence, four and a half to motivation, and only one page to drive seems an unjustifiable neglect of three very significant topics in the description of human behavior, particularly since sixty pages are devoted to a generally commendable treatment of the sensory processes.

The author has included in the first chapter a section on "Determinism in Psychology." His purpose in doing so he states in the preface as follows, "The writer feels that the issue might as well be raised at the very outset. The average student comes to psychology with a decidedly anti-deterministic point of view, full of free-will and mystical notions regarding the mental life. Only when the student gets clearly the idea that events of mind have their natural antecedents and conditioning factors just as is the case in physics and chemistry is he ready to proceed with the course." One may question whether any student will be so easily rid of habits of thinking anti-deterministically about human nature. On the other hand, let us not suggest to him that psychology need be on the defensive. Not any discussion of psychological determinism at the outset will accomplish a change in his attitude, but rather the content of the course itself, and incidentally of its textbook. Dr. Skaggs' book is a thoroughgoing scientific treatment and needs no such "defense" in its introductory chapter.

CARLETON F. SCOFIELD.
University of Buffalo.

*Hommage au Dr. Ovide Decroly.* Brussells, Belgium: Privately printed, 1934, pp. 512.

Some great men are so modest that they do their work and pass on before any but a few favored ones have even heard of them. Some useful books are known and appreciated by only a few readers. Here we have a book and the man whom it is intended to honor, of whom the foregoing statement is eminently true.

A few friends planned to celebrate the sixtieth birthday of Dr. Ovide Decroly and the twenty fifth anniversary of the founding of his famous school, L'Hermitage. Before the honorary volume and the other arrangements were completed, joy was turned to sorrow by the sudden death of Dr. Decroly. The "Festschrift" became a memorial

volume. A remarkable volume it is. Of its sixty one contributors, twenty-three have written of Dr. Decroly and his work. Thirty-eight have contributed original papers on educational topics.

The large number of nationalities represented is not only a rare tribute to Dr. Decroly and his genius, but it renders the volume of unusual value to educators the world over. It is not often that the views of such representative men from widely divergent nationalities are obtainable. The countries represented, include besides Belgium, France, Germany, England and America the following: Switzerland, Luxenbourg, Austria, Spain, Portugal, Holland, Hungary, Sweden, Rumania, Bulgaria, Italy, Poland, Finland, Czechoslovakia, Columbia S.A., Japan.

The list of contributors includes such names as Claparéde, Geheeb, Noguchi, Bovet, Burt, Demoor, Piaget, Pieron, Overstreet, Pintner and Zilliacus. All articles not in English, French or German have a satisfactory resumé in French.

It is a book that every educator will want and it is to be hoped that it will be available in college and public libraries. Being privately printed in Belgium, it is not quite so easy to secure, but undoubtedly can be procured by American booksellers who specialize in foreign publications.

HENRY H. GODDARD.

Ohio State University.